

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2003, 13:24:48 ; Search time 3691 Seconds
(without alignments)
10529.435 Million cell updates/sec

Title: US-10-053-410-3

Perfect score: 950

Sequence: 1 aaaaaaacccctcgcgat.....aaaaaaaaaaaaaaaaaaaaa 950

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 2889711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sta.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sta.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pin.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sv.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	932	98.1	932	8	AF371278	AF371278 Zea mays
2	171.8	18.1	757	8	OS19KDGLO	X63990 O.sativa mr
3	171.8	18.1	825	8	RICMSF	L12352 Oryza sativ
4	170.2	17.9	1773	8	D50643	D50643 Oryza sativ
5	170.2	17.9	99538	2	AC113332	AC113332 Oryza sat
6	170.2	17.9	124132	2	AC130605	AC130605 Oryza sat
7	158.2	16.7	102842	8	AF497474	AF497474 Aegilops
8	156.8	16.5	120562	8	AV268139	AV268139 Hordeum v
9	91.2	9.6	1859	8	HVDNAHOR3	X84368 H.vulgare H
10	91.2	9.6	2296	8	BLYHOR3	D82941 Barley Hor3
11	91.2	9.6	2380	6	BD012675	BD012675 Method of
12	91.2	9.6	2434	6	BD012674	BD012674 Method of
13	80	8.4	2804	8	AY248704	AY248704 Aegilops
14	80	8.4	2809	8	ASU39229	U39229 Aegilops ta
15	80	8.4	3095	8	TAGLUUDG	X03041 Wheat gene
16	79.8	8.4	101083	2	AC139010	AC139010 Homo sapi
17	78.4	8.3	6462	8	TAGLD12B	X12929 Triticum ae
18	78.2	8.2	537	8	AF226698	AF226698 Aegilops
19	78	8.2	1494	8	AY299518	AY299518 Thinopyru
20	77.6	8.2	553	8	TAE308970	AJ308970 Triticum
21	77.6	8.2	598	8	TAE308967	AJ308967 Triticum
22	76.4	8.0	153180	2	AP004683	AP004683 Oryza sat
23	75.8	8.0	1905	8	AF476960	AF476960 Aegilops
24	75.2	7.9	906	8	AY263345	AY263345 Thinopyru
25	75	7.9	1980	8	AY263343	AY263343 Thinopyru
26	75	7.9	1998	8	ACY306973	AJ306973 Aegilops
27	74.8	7.9	1494	8	AY264065	AY264065 Thinopyru
28	74.4	7.8	1962	8	ACY306974	AJ306974 Aegilops
29	73.8	7.8	1052	11	PM2H12B	AL684839 Penicilli
30	73.6	7.7	2330	8	TAE314785	AJ314785 Triticum
31	73.6	7.7	2409	8	AF513640	AF513640 Aegilops
32	73.6	7.7	2996	8	X61026	X61026 Wheat Glu-1
33	73.4	7.7	1971	8	AF476962	AF476962 Aegilops
34	73.4	7.7	2163	8	AY245797	AY245797 Triticum
35	72.4	7.6	186752	9	AC111200	AC111200 Homo sapi
36	72	7.6	2331	8	AF216869	AF216869 Triticum
37	72	7.6	2331	8	SCE314781	AJ314781 Secale ce
38	71.8	7.6	1827	8	AY260548	AY260548 Triticum
39	71.8	7.6	1827	8	AY260549	AY260549 Triticum
40	71.8	7.6	1881	8	AY174159	AY174159 Aegilops
41	70.4	7.4	2297	8	AY249141	AY249141 Triticum
42	70.4	7.4	2331	8	SCE314774	AJ314774 Secale ce
43	70.4	7.4	2331	8	SCE314775	AJ314775 Secale ce
44	70.4	7.4	2331	8	SCE314777	AJ314777 Secale ce
45	70.4	7.4	2331	8	TAE314783	AJ314783 Triticum

ALIGNMENTS

RESULT 1
AF371278
LOCUS
DEFINITION Zea mays alpha globulin mRNA, complete cds.
ACCESSION AF371278
VERSION AF371278.1 GI:16305141
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 932)
AUTHORS Woo,Y.M., Hu,D.W., Larkins,B.A. and Jung,R.
TITLE Genomics analysis of genes expressed in maize endosperm identifies

JOURNAL	novel seed proteins and clarifies patterns of zein gene expression
MEDLINE	Plant Cell 13 (10), 2297-2317 (2001)
PUBMED	21480069
REFERENCE	11595803
AUTHORS	2 (bases 1 to 932)
TITLE	Jung,R.
JOURNAL	Direct Submission
FEATURES	Submitted (18-APR-2001) TTD, Pioneer Hi-Bred International, Inc., 7300NW 62nd Ave., Johnston, IA 50131, USA
source	Location/Qualifiers
	1..932
	/organism="Zea mays"
	/mol_type="mRNA"
	/variety="B73"
	/db_xref="taxon:4577"
	/tissue_type="endosperm"
CDS	108..728
	/codon_start=1
	/product="alpha globulin"
	/protein_id="AA116993.1"
	/db_xref="GI:16305142"
	/translation="MAKIAAAAAALCAFAALVAVVCQGEVERQLRDLCQWQEVQSS PLDACQVLRLQLTGGGGGGVGFPRWGTGLRMCCQQLQDVSRCECAIRSMVRGY EEMPPLEKXGWPMRQGPQPPQGGGGGGGYIYPCSRPGEYGYGGGQRMYPPCR PGTGGGPRGRLVTLKARBYAAGLPMCLSLPQECISFGGDQY"
BASE COUNT	199 a 259 c 334 g 140 t
ORIGIN	
Query Match	98.1%; Score 932; DB 8; Length 932;
Best Local Similarity	100.0%; Pred. No. 1.1e-138;
Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	4 AAAACCCCTCGTCGATCACCAACAAAGAACACAGTAACCTAGCAGCTAGCACATCAACA 63
Db	1 AAAACCCCTCGTCGATCACCAACAAAGAACACAGTAACCTAGCAGCTAGCACATCAACA 60
Qy	64 AGTGGCAGACAGACAAAGATTGTGAGGGTGATCCGCGCTGAGAAAGATGGCTAAGATCG 123
Db	61 AGTGGCAGACAGACAAAGATTGTGAGGGTGATCCGCGCTGAGAAAGATGGCTAAGATCG 120
Qy	124 CCGCGGCGGCGCGCGCGCTGTGCTTTCGGGGCCCTGGTGGCCGTGGCCGTCTGCCAAG 183
Db	121 CCGCGGCGGCGGCGGCGCGCTGTGCTTTCGCGCCCTGGTGGCCGTGGCCGTCTGCCAAG 180
Qy	184 GCGAGTGTGAGCGCAGAGGCTCAGGACCTGCACTGTGCAAGAGGTCCAGGAGAGCC 243
Db	181 GCGAGTGTGAGCGCAGAGGCTCAGGACCTGCACTGTGCAAGAGGTCCAGGAGAGCC 240
Qy	244 CGTTCGACGGGTGCGGCAGGTCTCTGACCGGCAGCTAACCGGCGGCGGCGGCGCGCG 303
Db	241 CGTTCGACGGGTGCGGCAGGTCTCTGACCGGCAGCTAACCGGCGGCGGCGGCGCGCG 300
Qy	304 GCGTTGCCCGCTTCGCGTGGGGACCGCGGCTCCGGATCCGGTGTGTCGACGAGTCCAGG 363
Db	301 GCGTTGCCCGCTTCGCGTGGGGACCGCGGCTCCGGATCCGGTGTGTCGACGAGTCCAGG 360
Qy	364 ACGTGACCCGCGAGTGCCTGCGCCGCATCCGAGCATGCTCAGGGGCTACGAGGAGG 423
Db	361 ACGTGACCCGCGAGTGCCTGCGCCGCATCCGAGCATGCTCAGGGGCTACGAGGAGG 420
Qy	424 CCATGCCCGCGCTGGAGAAAGGCTGTGTGGCATGGGGGCGGCAGCAGCGCGCGCGCG 483
Db	421 CCATGCCCGCGCTGGAGAAAGGCTGTGTGGCATGGGGGCGGCAGCAGCGCGCGCGCG 480
Qy	484 AGGAGAGGAGGAGGAGGACAGGGGGGCTACTACTACCCCTGACGCCGCGCAGAGAGGAT 543
Db	481 AGGAGAGGAGGAGGAGGACAGGGGGGCTACTACTACCCCTGACGCCGCGCAGAGAGGAT 540
Qy	544 ACGCTACCGTCAAGGTGGCCAGCGGAGATGTATCCACGTTGCTGTCGGGACACACCG 603
Db	541 ACGCTACCGTCAAGGTGGCCAGCGGAGATGTATCCACGTTGCTGTCGGGACACACCG 600
Qy	604 GCGCGGCGGCGCAAGGATCGCGCGCTGAGGCTTACGAAGGGCCCGGAGATGAGCCCGGGT 663

```
terminator 681.688
polya_signal 707.712
terminator 721.728
polya_site 742
BASE COUNT 157 a 197 c 263 g 140 t
ORIGIN
/translation="MAAMVAISGAHVSRSEMRFRDQCQREVDSPDLACROVLDRQL
TGRERFQPMFRPGALGRMCCQLODVSRCCAAIRRMVRSYEESSMPLEQWS
SSSEYVGGEGSSSQGYGEGSSEGYEQOQPGMTRVLRARQVAAQLPSMCR
VEPQCSIFAAQY"
Query Match 18.1%; Score 171.8; DB 8; Length 757;
Best Local Similarity 71.6%; Pred. No. 8e-18;
Matches 240; Conservative 0; Mismatches 92; Indels 3; Gaps 1;
QY 138 GCGGCGCTGTGTTCCGGCCCTCGTGGCGCGCTGCGCCGCTGCAAGCGCAGGTCGAGCGG 197
DB 53 GCGGCGCGCTCATGGCGCCCATGTGGCCATCTCCGGCGCGCACGTGACGAGTCGGAG 112
QY 198 CAGAGCTCAGGACCTGCGAGTCTGGCAGGAGGTCCAGGAGAGCCCGCTCGACGGGTGC 257
DB 113 ATGAGTTTCAGGACAGGAGTGCAGCGGAGGTGTCAGACAGCCCGCTGACACGGTGC 172
QY 258 CGCCAGGTCTCGACCGGAGCTAACCGGCGCGCGCGCGCGCGCTTGCCCGCTTC 317
DB 173 CGGAGGTGCTCGACCGGAGCTCACCGCGCGGAGAGTTCAGCCGATGTTCCGCCGC 232
QY 318 CGGTGGGACACCGGCTCCGATGCGGTGCTGCGCAGCAGCTCCAGACGTGAGCCCGAG 377
DB 233 CGGGCGCGCTCGGCTCGCGATGTCAGTGTGTCAGCAGCTGCGAGACGTGAGCCCGAG 292
QY 378 TGCCGCTGCGCGCCATCCGAGCATGTGTCAGGGCTACGAGAGGCCATGCC---GCCG 434
DB 293 TGCCGCTGCGCGCCATCCGCGGATGGTGAGGAGCTACGAGAGAGCATGCCGATGCC 352
QY 435 CTGGAGAAAGGTGTGGCCATCGGCGCGCGCGAGCA 469
DB 353 CTGGAGAAAGGTGTGGCTCGTCTCGTGAGTA 387
RESULT 3
LOCUS Oryza sativa DNA fragment with a miscellaneous signal and an open
DEFINITION reading frame.
ACCESSION L12252
VERSION L12252.1 GI:169804
KEYWORDS
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE 1 (bases 1 to 825)
AUTHORS Kriehnan, H.B. and Pueppke, S.G.
TITLE Nucleotide sequence of an abundant rice seed globulin: homology
with the high molecular weight glutelins of wheat, rye and
triticale
JOURNAL Biochem. Biophys. Res. Commun. 193 (1), 460-466 (1993)
MEDLINE 93277591
PUBMED 8503935
COMMENT Original source text: Oryza sativa (cultivar Lamont) 20 days
post-anthesis endosperm cDNA to mRNA.
FEATURES
source
1..825
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Lamont"
/db_xref="taxon:39947"
/tissue_type="endosperm"
/dev_stage="20 days post-anthesis"
1..54
misc_signal
25..549
/codon_start=1
/label=ORF
/protein_id="AAA72362.1"
/db_xref="GI:169805"
```

```
/translation="MAAMVAISGAHVSRSEMRFRDQCQREVDSPDLACROVLDRQL
TGRERFQPMFRPGALGRMCCQLODVSRCCAAIRRMVRSYEESSMPLEQWS
SSSEYVGGEGSSSQGYGEGSSEGYEQOQPGMTRVLRARQVAAQLPSMCR
VEPQCSIFAAQY"
BASE COUNT 165 a 244 c 270 g 146 t
ORIGIN
Query Match 18.1%; Score 171.8; DB 8; Length 825;
Best Local Similarity 71.6%; Pred. No. 8e-18;
Matches 240; Conservative 0; Mismatches 92; Indels 3; Gaps 1;
QY 138 GCGGCGCTGTGTTCCGGCCCTCGTGGCGCGCTGCGCCGCTGCAAGCGCAGGTCGAGCGG 197
DB 13 GCGGCGCGCTCATGGCGCCCATGTGGCCATCTCCGGCGCGCACGTGACGAGTCGGAG 72
QY 198 CAGAGCTCAGGACCTGCGAGTCTGGCAGGAGGTCCAGGAGAGCCCGCTCGACGGGTGC 257
DB 73 ATGAGTTTCAGGACAGGAGTGCAGCGGAGGTGTCAGACAGCCCGCTGACACGGTGC 132
QY 258 CGCCAGGTCTCGACCGGAGCTAACCGGCGCGCGCGCGCGCTTGCCCGCTTC 317
DB 133 CGGAGGTGCTCGACCGGAGCTCACCGCGCGGAGAGTTCAGCCGATGTTCCGCCGC 192
QY 318 CGGTGGGACACCGGCTCCGATGCGGTGCTGCGCAGCAGCTCCAGACGTGAGCCCGAG 377
DB 193 CGGGCGCGCTCGGCTCGCGATGTCAGTGTGTCAGCAGCTGCGAGACGTGAGCCCGAG 252
QY 378 TGCCGCTGCGCGCCATCCGAGCATGTGTCAGGGCTACGAGAGGCCATGCC---GCCG 434
DB 253 TGCCGCTGCGCGCCATCCGCGGATGGTGAGGAGCTACGAGAGAGCATGCCGATGCC 312
QY 435 CTGGAGAAAGGTGTGGCCATCGGCGCGCGCGAGCA 469
DB 313 CTGGAGAAAGGTGTGGCTCGTCTCGTGAGTA 347
RESULT 4
LOCUS Oryza sativa (japonica cultivar-group) DNA for 26 kDa globulin,
DEFINITION complete cds.
ACCESSION D50643
VERSION D50643.1 GI:840704
KEYWORDS 26 kDa globulin.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE 1 (bases 1 to 825)
AUTHORS Nakase, M., Hotta, H., Adachi, T., Aoki, N., Nakamura, R., Masumura, T., Tanaka, K. and Matsuoka, T.
TITLE Cloning of the rice seed alpha-globulin-encoding gene: sequence
similarity of the 5'-flanking region to those of the genes encoding
wheat high-molecular-weight glutenin and barley D hordein
JOURNAL Gene 170 (2), 223-226 (1996)
MEDLINE 96235139
PUBMED 8666249
REFERENCE 2 (bases 1 to 1773)
AUTHORS Nakase, M., Hotta, H., Adachi, T., Alvarez, A.M., Aoki, N., Nakamura, R., Masumura, T., Tanaka, K. and Matsuoka, T.
TITLE Gene cloning and characterization of rice seed 26 kDa globulin:
sequence similarity of its 5'-flanking region to those of wheat
high molecular weight glutenin and barley D hordein
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1773)
AUTHORS Nakase, M.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-1995) Masayuki Nakase, Nagoya University, School
of Agricultural Sciences, Applied Biological Sciences; Furo-cho,
Chikusa-ku, Nagoya, Aichi 464-01, Japan
(E-mail: i45231a@nuc.cc.nagoya-u.ac.jp, Tel: 052-789-4131,
Fax: 052-789-4120)
```


mRNA	375	GAGTCCCGCTGCGCCGCATCCGAGCATGTCTACGGGCTTACGAGGAGGCCATCCCGCCG 434
Db	3578	GAGTCCCGCTGCGCCGCCTCCGGGATGCTGCGGGACTACGAGCAGTCCATCCCGCCG 3637
CDS	435	CTGGAGAAAGCTGTGCGCCATGGGCGGCAGCAGCGCCGCCCGCAGGG 487
Db	3638	CTGGGAGAAGGCGCCACCGCTCGTGGGGAGCGTCAACAAGAGCAGGGGTG 3690
RESULT 8		
LOCUS	AY268139	120562 bp DNA linear PLN 07-MAY-2003
DEFINITION	Hordeum vulgare BAC 184G9, complete sequence.	
ACCESSION	AY268139	
VERSION	AY268139.1	GI:30421164
KEYWORDS	Hordeum vulgare	
SOURCE	Hordeum vulgare	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Hordeum.	
REFERENCE	1 (bases 1 to 120562)	Gu, X.Q., Anderson, O.D., Londeore, C.F., Kong, X., Chibbar, R.N. and Lazo, G.R.
AUTHORS	Structural organization of the barley D-hordein locus in comparison with its orthologous regions of wheat genomes	
TITLE	Unpublished	
JOURNAL	2 (bases 1 to 120562)	Gu, X.Q., Anderson, O.D., Londeore, C.F., Kong, X., Chibbar, R.N. and Lazo, G.R.
REFERENCE	Direct Submission	
AUTHORS	Submitted (03-APR-2003) ARS, U.S. Department of Agriculture, 800 Buchanan Street, Albany, CA 94710, USA	
JOURNAL	Location/Qualifiers	
FEATURES	1..120562	
source	/organism="Hordeum vulgare"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:4513"	
	/chromosome="5"	
	/clone="BAC 184G9"	
	/clone_lib="HV_Mba-Library, Clemson University of Genomics Institute"	
	complement(725..3761)	
repeat_region	/transposon="LTR retrotransposon Usier_184G9-1"	
	complement(3770..12377)	
repeat_region	/transposon="LTR retrotransposon Bare-1_184G9-1"	
	complement(12383..21304)	
repeat_region	/transposon="LTR retrotransposon Usier_184G9-1"	
	<26067..>26245	
repeat_region	/note="Afai repeat"	
	<26271..>26538	
repeat_region	/note="Afai repeat"	
	27437..27600	
repeat_region	/transposon="MITE Stowaway"	
	complement(join(<27857..28230,28919..>31616))	
mRNA	/product="putative receptor kinase"	
	complement(join(27857..28230,28919..31616))	
CDS	/note="similar to rice XA21"	
	/codon_start=1	
	/product="putative receptor kinase"	
	/protein_id="BAP31049.1"	
	/db_xref="GI:30421165"	
	/translation="METRPMRLRLIALITTTTLLHPSTSSSVSTAHDLPLLSPFSKVS LITKPLDGLSSWTKTNGSTGFCNMTGVECSAHPGHKALRQLGSLGTISPTFLN LSRLRALDGLSKLQGGQIPSSIGCNCFALRTLNLSVLSGAIIPAMGNLSKLVLSVPS KNDISGTPSTFAGLATVAVFSVARNHVGPWMLGNLTALDNLNADIMSGHVPV ALSKLNLSRTVAIINNLOGLIPVLFNWMSLECLNFGSNLSGLPQD1GSMPLNKL KFSVYFNRFQIPASLNISLSLEHLNHRFRGRIPSNIQSGRLTVFVGNNEQLQ ATESRDWFLTLANCLLNLIQNLNLSGILNPNSIGNLSKLSQKLEURVGVGNQAGLI PTGIRYKLAILEAFADNRFTGTIPSDIGKLSNLKELSLFNQRYIGPISIPGFLNGL NLALSTNNLNGSLPATPGNTLISLNLSSGLIPBEVVMRISLALFLNLNNL LDGTPSHIGOLANLAIIDFNSKLSGIPINALGSCIALOFLHOGNLLQGOIPKELM NLLGTSPIHIGOLANLAIIDFNSKLSGIPINALGSCIALOFLHOGNLLQGOIPKELM NLLGTSPIHIGOLANLAIIDFNSKLSGIPINALGSCIALOFLHOGNLLQGOIPKELM	

```
ALRGLEELDLSNNLSGVPPEFLESFOLLKNLNLNLSFNHLSGVPVDPKGI FSNASVISLT
SGNGLQGGVFFHPTCPYSPDPKLAHSHKLLQILVFTAVGAFILGLVCITACRQYWNKS
RNDHMQDQENIPEMFOTSYTELHATDSFSEENLVGRSGVSGVKGTSGGANLITA
AVKLDVORQAGATRSFISECNALKWRHKLKVLTVCDSLHDHSGNOFKALVLEFIPN
GSLDKWLPSTDEDETRFENLMQRNLIALDVAELKIDHIDPPIVCHDVRKPSNILLD
DMVAHLGDFGLAKIIRAKSKSLADQSCSVIGIYVAPEYGTETGIESVEGDVY
SYGVLLLEMLLTCRRRTDFSTTNLPKLVEMACFNLENTMDVNRINCNEQPAVLEL
FAAPVSRGLACRQSARQIRKMGDVVKELGAIKQIIMASQNYASWSKLY"
31751..31774
/rpt_type=tandem
/rpt_unit=ttc
38555..38618
/transposon="MITE"
<38912..>39586
/product="globulin"
38912..39586
/codon_start=1
/product="globulin"
/protein_id="AAP31050.1"
/db_xref="GI:30421166"
/translation="MGKTFPFAVFTLTITISAAQGVLEQSLADACRCRGEVQAKPLLA
CRQILEHOLTGRAVGRPFAQMGARDRCQOLESVSRGCSALRGVWRDYEQSMPP
LREGRRSSGERQQQGGSGESTAEQQEVQGGQYGTGSGEQQQGGYGHVTVGRG
GQQQGMCLREPRQKQGGGSGSGAQKPKVGRVRLTKVRLPTACRIEPEQCSVFS
TLPVLG"
complement(41245..42179)
/transposon="Solo-LTR retrotransposon Bare-1_184G9-2s"
complement(42180..42386)
/transposon="LTR retrotransposon Lolao9_184G9-1p"
complement(42387..>43152)
/transposon="LTR retrotransposon Bare-1_184G9-3p"
complement(43153..53697)
/transposon="LTR retrotransposon Lolao9_184G9-1p"
complement(53698..53841)
/transposon="Solo-LTR retrotransposon Bare-1_184G9-2s"
53842..58439
/transposon="LTR retrotransposon Leo9yg_184G9-1"
complement(58440..59149)
/transposon="Solo-LTR retrotransposon Bare-1_184G9-2s"
<62032..>64305
/product="D-Hordein"
62032..64305
/codon_start=1
/product="D-Hordein"
/protein_id="AAP31051.1"
/db_xref="GI:30421167"
/translation="WAKRLVLFVAVIVAVLVALTAREINGNNIFLDSRSRLQOCERE
LOESLEACRRVVDQVLQGLPWSLQWCCQQLRDVSPRCVVALSVQVRYQQQT
EYPSKGSGYPGPTAPPLOQGGWGTSTVKWYIPDOTSQSQSWQGGYHVSQTSQQT
GGQGGSYPGSTFPQPGQGGQPGQRPWSPATFPQPGQGGQGGYYPGATSLIQ
PGQGGQGPYQSATSPOPGQGGQGGQEPYPIATSPHQRGQGGQGGYVPSVTSPO
QSGQGGQYPTSTSPQSGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
SQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
HQQQLVQQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
SQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
SGSQVQGAQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
LHQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
TTVSPHQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
QQTTVSPHQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
QQQQPQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
LPAMCRLEGGGGLASQ"
join(<64782..64944,65077..65414,65452..65555,66035..66224,
66334..66725,66880..>67194)
/product="putative protein kinase"
join(64782..64942,65077..65414,65452..65555,66035..66224,
66334..66725,66880..67194)
/codon_start=1
/product="putative protein kinase"
/protein_id="AAP31052.1"
/db_xref="GI:30421168"
/translation="MIGTRVLAHSRVDPAIRGVAAARGRVVFAAIRWGAARGRVVFA
AVRWGAARGTRKREAGTSTSETKKTKRWGRFGCGMASHEVEPLTSETKKTKRW
GRGFCGMASHEAEPLTSTKKRKNVAASBPDKRPFNKIWKKKAKNQALATLV
KEISLATKLSAMHVNINLSMNICPTQTVEHSGLYLRNLAVIAVKQLDKDGLQGNRE
```

```
FLVEVLMLSLHHNPNLVTLLGYCTECDOQKILVYEMPLGSLQDHLDDLTPKSPQLSWH
TRMKIADVAYLEHVPVPPVDRDKASNILLDGNFSAKLADFLGAKLGPVGDK
THVTRVMCTYGYCAPEYAMGPKLTKMSDIYCFVGLLELITGRRADITTKPTRQIL
VHMAAPLKDKKKTMMADPLDKSKYPLKGLYQALAISSMCIQREAI SRPLISDVVTA
LTFIADPNYDPPDIEPLPISVPNYDKGISREAEISISGPEEKQVEUS"
70339..73731
/transposon="non-LTR retrotransposon Leva_184G9-1"
70469..70482
/rpt_type=tandem
/rpt_unit=ta
complement(77378..>77855)
/transposon="LTR element Wobi_184G9-1p"
77856..86572
/transposon="MITE"
complement(<86571..87579)
/transposon="LTR retrotransposon Bare-1_184G9-4"
87266..87285
/transposon="LTR element Wobi_184G9-1p"
87286..87285
/rpt_type=tandem
/rpt_unit=ag
complement(87580..89305)
/transposon="LTR retrotransposon solo-LTR Bare-1_184G9-5s"
complement(<89306..90807)
/transposon="LTR element Wobi_184G9-1p"
complement(<91388..>91975)
/transposon="LTR retrotransposon Wham_184G9-1p"
complement(<93655..>95251)
/transposon="LTR retrotransposon Barbara_184G9-1p"
95284..>98782
/transposon="LTR retrotransposon Usier_184G9-2p"
complement(<98783..100481)
/transposon="LTR retrotransposon Bare-1_184G9-6p"
100482..101938
/transposon="LTR retrotransposon solo-LTR Wham_184G9-2s"
101839..109132
/transposon="LTR element Egug_184G9-1"
109133..109156
/transposon="LTR retrotransposon solo-LTR Wham_184G9-2s"
complement(109157..>109388)
/transposon="LTR retrotransposon Bare-1_184G9-6p"
complement(109389..119184)
/transposon="LTR retrotransposon Wham_184G9-3"
complement(<119185..120562)
/transposon="LTR retrotransposon Bare-1_184G9-6p"
BASE COUNT 32679 a 27422 c 28340 g 32121 t
ORIGIN
Query Match 16.5%; Score 156.8; DB 8; Length 120562;
Best Local Similarity 71.3%; Pred. No. 1.8e-15;
Matches 238; Conservative 0; Mismatches 87; Indels 9; Gaps 2;
QY 141 GGCTGTGCTTCGCGGCCCTGTGTGGCGCGTCTGCAAGGCGAGGTCCAGGCGGCGGCGAG 200
Db 38933 GCGGTGTTCCTGACACCTGTGTGACCATCTCCGCGGCCAAGGCGGTGCTCGA--GCAG 38989
QY 201 AGGCTCAGGAGCCTTCAGTGTCTGGCAGGAGGTCTCAGGAGAGCCGCTCCACCGTCCGCG 260
Db 38990 AGCTTCGCGAGCGCAGTGTCCGCGGCGAGGTCTCAGGCGAAGCCGCTCTCGGTGCGCG 39049
QY 261 CAGGTCTTCGACCGGCGAGCTAAACCGCGCGCGCGCGCGCGCGGTTCGGCCGCTTCGCG 320
Db 39050 CAGATCTTCGAGCATCAGTGTACCGCGCGCGCGCGCGCGGTTCAGGCGCGCGAG 39109
QY 321 TGGGCGACCGGCTCCGGATGCGGTGCTGCCAGCAGCTCCAGAGCGGAGCCATGCGCGCTGGAG 380
Db 39110 TGGGCGCGCAGGAC-----CGGTGCTGCCAGCAGCTCAGAGCGGTGAGCGCGGCTGC 39163
QY 381 CGCTGCGCGCCATCCGAGCATGTGTACGGGCGCTACGAGGAGCCATGCGCGCTGGAG 440
Db 39164 CGCTGCTCGCCCTCGTGGCATGTGTGGGAGTACGAGCAGTCCATGCGCGCGCTGAGG 39223
QY 441 AAAGCTGTGTGCCATGGGGCGCGGAGCAGCAGC 474
Db 39224 GAAGGCGCGCGCGCTCGTTCGGGGGAGCGTCAAC 39257
```

Matches 178; Conservative 0; Mismatches 88; Indels 24; Gaps 1

Qy 207 AGGGAACCTGTCAGTGTGGCAGGAGGTCCAGGAGAGCCCGCTCGACGCGTGCAGGAGGTC 266

Db 543 AGGCAGTACAGTGTGAGCGCGAGCTCCAGGAGAGCTCGCTCGAGGCGTGCAGGCGGGTC 602

Qy 267 CTCGACGGGAGATTACCGGCGGCGGCGGCGGCGGCGTGTGGCCCGTTCCGGTGGGCGC 326

Db 603 GTGAGCAACAGCTG-----GTTGGCCAGCTGCCATGGAGC 638

Qy 327 ACCGGGCTCGGATGCGGTGCTGCCAGAGCTCCAGAGCGTGAAGCGGAGTCCCGCTGC 386

Db 639 ACCGGGCTCAGATGCGAGTGTCTCCAGAGCTTCGGAGCGTCAAGCGGAGTCCCGCTGC 698

Qy 387 GCGCCCATCGGAGCATGTGTACGGGGCTACGAGGAGGCCCATGCGCGCTGGAGAAAGGC 446

Db 699 GTCCGCTCAGCGAGTCTGTAGGCAATACGAGACCAACCGAGGTGCCATCAAGGGA 758

Qy 447 TGTGTGCGCATGGGGGCGGCGAGCAGCAGCGCCCGCGGAGGAGGAGGAGG 496

Db 759 GGATCTCTTACCCGCGGCGGACCGCACGCGCGCTGCAGCAGAGGAGATG 808

RESULT 10

BLYHOR3
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BLYHOR3 2296 bp mRNA linear PLN 06-FEB-1999
Barley Hor3 mRNA for D hordein, complete cds.

D82941
D82941.1 GI:1167497
D hordein; Hor3.
Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 2296)
Hirota,N., Kuroda,H. and Ito,K.
Unpublished
2 (bases 1 to 2296)
Hirota,N.
Direct Submission
Submitted (06-JAN-1996) Naohiko Hirota, Plant Bioengineering
Research Laboratories,Sapporo breweries, Biotechnology department;
Kizaki 37-1, Nitta, Gunma 370-03, Japan
(E-mail: sapplant@po.infosphere.or.jp, Tel:0276-56-1455,
Fax:0276-56-1605)

FEATURES
Location/Qualifiers

1..2296
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/strain="Haruna Nijo"
/db_xref="taxon:112509"
/clone="DH4"
/tissue_type="Endosperm"
/dev_stage="Seed"
1..2296
/gene="Hor3"
37..2160
/gene="Hor3"
/function="seed storage"
/codon_start=1
/product="D hordein"
/protein_id="BAA11642.1"
/db_xref="GI:1167498"
/translation="MAKRLVLFAVIVALTAAIREINGNIFLDSRSRLQOCERE
LQESLEACRRVVDQLVGLPMTGLQCCQLRDSPECRVALSVQVRYEQQT
EYPSGGSPYGTAPPLQQGGWGTGSKVYYPDQTSSQQSGQQGQYHSVTSSQ
GGQGGSPYGTFTPQQGGQGPQRQWSPYSATFPQPGGQGGQGYPGATSLQ
PGQGGQPYQSATSPQPGQGGQDETYPATSPHQPGWQPGQGGQGYFYSVTS
SQGGQGGQYPTTSQPGQGGQGLGQGGQGGQYFYSATFPQPGWQGGQGSYST
SPQGGQGGQYNTSGTSTQPGQGGQGGQGYPIATSPQPGQGGQGLGQGGQ
HQQGLVQGGQGGQGGQYPSMTSPHTQGGQGYPSAISPGQGGQGGQYPSGAS
SGQGLYACQHSSTSPQQAQCSASSPQGLGLYPSGAYTQPKPGQGYNPGTGS

gene
CDS

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2003, 13:28:03 ; Search time 2415 Seconds

(without alignments)

9560.767 Million cell updates/sec

Title: US-10-053-410-3

Perfect score: 950

Sequence: 1 aaaaaaacccctcgtcat.....aaaaaaaaaaaaaaaaaaaa 950

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_esti:*

9: gb_est1:*

10: gb_est2:*

11: gb_est3:*

12: gb_est4:*

13: gb_est5:*

14: gb_est6:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pin:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pig:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
C	1	894.2	94.1	935	29	CC349094
	2	893	94.0	900	11	AY105070
	3	732	77.1	1038	29	CC349105
C	4	648.6	68.3	674	14	CA401604

10000	C	5	520.2	54.8	570	14	CA402183	CA402183	EL01N0432
10000	C	6	497	52.3	602	9	AI711821	AI711821	605064E02
10000	C	7	456.4	48.0	539	9	AI745997	AI745997	605078E03
10000	C	8	430.6	45.3	560	9	AI712201	AI712201	605065F05
10000	C	9	385.2	40.5	473	14	CD443222	CD443222	EL01N0432
10000	C	10	374	39.4	586	10	BE917827	BE917827	OVI_7_F02
10000	C	11	362.6	38.2	605	10	BE917909	BE917909	OVI_7_F02
10000	C	12	340.2	35.8	367	12	BM500544	BM500544	PAC000000
10000	C	13	307.2	32.3	389	9	AI670620	AI670620	605035H03
10000	C	14	285	30.0	652	14	CA402267	CA402267	EL01N0434
10000	C	15	256.2	27.0	425	10	EG048804	EG048804	OVI_23_CO
10000	C	16	209.8	22.1	496	10	EG049097	EG049097	OVI_23_CO
10000	C	17	170.2	17.9	443	9	AU182658	AU182658	AU182650
10000	C	18	170.2	17.9	453	9	AU182650	AU182650	AU182650
10000	C	19	170.2	17.9	522	9	AU076088	AU076088	AU076088
10000	C	20	170.2	17.9	537	13	BQ172833	BQ172833	1011b04
10000	C	21	170.2	17.9	595	13	BQ172799	BQ172799	1008b04
10000	C	22	168.8	17.8	444	14	C20089	C20089	C20089
10000	C	23	167.8	17.7	460	14	C20009	C20009	C20009
10000	C	24	167.8	17.7	461	14	C20289	C20289	C20289
10000	C	25	167.6	17.6	712	9	AU163963	AU163963	AU163963
10000	C	26	166.8	17.6	704	9	AU094496	AU094496	AU094496
10000	C	27	165.4	17.4	466	9	AU182628	AU182628	AU182628
10000	C	28	165.2	17.4	459	14	C20342	C20342	C20342
10000	C	29	164.4	17.3	452	14	C20476	C20476	C20476
10000	C	30	163.2	17.2	549	9	AA751601	AA751601	96AS0337
10000	C	31	163	17.2	687	9	AU163860	AU163860	AU163860
10000	C	32	161.8	17.0	683	9	AU164031	AU164031	AU164031
10000	C	33	161.6	17.0	707	9	AU164041	AU164041	AU164041
10000	C	34	161.4	17.0	497	13	BQ247039	BQ247039	Tae15002B
10000	C	35	161.4	17.0	534	14	CA734375	CA734375	wde2f.pk0
10000	C	36	161.4	17.0	536	13	BQ246513	BQ246513	Tae15009D
10000	C	37	161.4	17.0	599	13	BQ246084	BQ246084	Tae15016E
10000	C	38	161.4	17.0	600	13	BQ246182	BQ246182	Tae15015B
10000	C	39	161.4	17.0	603	13	BQ246465	BQ246465	Tae15010D
10000	C	40	161.4	17.0	619	13	BQ241153	BQ241153	Tae15008E
10000	C	41	161.4	17.0	625	13	BQ806663	BQ806663	WHE3581
10000	C	42	160.6	16.9	476	14	CA734371	CA734371	wde2f.pk0
10000	C	43	159.8	16.8	547	9	AU165811	AU165811	AU165811
10000	C	44	159.6	16.8	393	13	BQ244740	BQ244740	Tae15038B
10000	C	45	159.6	16.8	440	13	BQ246456	BQ246456	Tae15010E

ALIGNMENTS

RESULT 1	CC349094/c	935 bp	DNA	linear	GSS 16-MAY-2003
LOCUS	OGUEN26TH ZM 0.7 1.5 KB	Zea mays	genomic clone	ZMMBMA0422E04,	
DEFINITION	genomic survey sequence.				
ACCESSION	CC349094				
VERSION	CC349094.1	GI:30818501			
KEYWORDS	GSS.				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.				
AUTHORS	1 (bases 1 to 935)				
TITLE	Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.				
JOURNAL	Consortium for Maize Genomics				
COMMENT	Unpublished				
	Contact: Cathy Whitelaw				
	TIGR				
	9712 Medical Center Drive, Rockville, MD 20850, USA				
	Tel: 301-838-5843				
	Fax: 301-838-0208				
	Email: whitelaw@tigr.org				
	Seq primer: TR				
	Class: sheared ends.				

```
FEATURES
source
  Location/Qualifiers
    1..935
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBMA0422E04"
      /clone_lib="ZM 0.7 1.5 KB"
      /notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
      methylation filtered genomic DNA library"
BASE COUNT    147 a    339 c    259 g    190 t
ORIGIN
Query Match      94.1%; Score 894.2; DB 29; Length 935;
Best Local Similarity 98.0%; Pred. No. 1.4e-80;
Matches 916; Conservative 0; Mismatches 18; Indels 1; Gaps 1;
QY 11 CCTCGTCGATCACCACCAAG-AACACAGTAACCTAGCAGCTAGCAGATCAACAAGTGGC 69
DB 935 CCTCGTCGATCACCACCAAGAACACAGTAACCTAGCAGCTAGCAGATCAACAAGTGGC 876
QY 70 GACAGACAAAGATTGTGAGGCTGATCCGCGCTGAGAGAGATGGCTAAGATCGCCGCG 129
DB 875 GACAGACAAAGATTGTGAGGCTGATCCGCGCTGAGAGAGATGGCTAAGATCGCCGCG 816
QY 130 CGCGCGCGCGGCTGTGCTTCCGCGCCCTGTGTGCGCGCTGTGCGCGCTGTGCGCGCG 189
DB 815 CGCGCGCGCGGCTGTGCTTCCGCGCCCTGTGTGCGCGCTGTGCGCGCTGTGCGCGCG 756
QY 190 TCGAGCGGCAGAGGCTCAGGAGCTGTGAGTGTGGCAGGAGTCCAGGAGAGCCGCTCG 249
DB 755 TCGAGCGGCAGAGGCTCAGGAGCTGTGAGTGTGGCAGGAGTCCAGGAGAGCCGCTCG 696
QY 250 ACAGGTCGCCAGGCTCTCGACCGGAGTAAACCGCGCGCGCGCGCGCGCGCGCGGTTG 309
DB 695 ACAGGTCGCCAGGCTCTCGACCGGAGTAAACCGCGCGCGCGCGCGCGCGCGCGGTTG 636
QY 310 GCCGCTCCGTTGGGCGACCGGCTCCGATGTGGTGTGCTCCAGCAGCTCCAGAGGTGA 369
DB 635 GCCGCTCCGTTGGGCGACCGGCTCCGATGTGGTGTGCTCCAGCAGCTCCAGAGGTGA 576
QY 370 GCCGCGAGTCCGCTGCGCGCCATCCGAGAGTGTGTCAGGAGTACGAGGAGCCATGC 429
DB 575 GCCGCGAGTCCGCTGCGCGCCATCCGAGAGTGTGTCAGGAGTACGAGGAGCCATGC 516
QY 430 CGCGCTGGAGAAAGCTGTGTGCCATGGGGCGGCGAGCAGCAGCGCGCGCGCGAG 489
DB 515 CGCGCTGGAGAAAGCTGTGTGCCATGGGGCGGCGAGCAGCAGCGCGCGCGAG 456
QY 490 GAGGAGGACAGAGGGGGTACTACTACCCCTGACCGCGCGCGCGCGAGAGGATACGGCT 549
DB 455 GAGGAGGACAGAGGGGGTACTACTACCCCTGACCGCGCGCGCGAGAGGATACGGCT 396
QY 550 ACAGTCAAGGTGCGCAGCGCAGATGTATCCACCGTGTGCTCCGCGCACCACCGCGCG 609
DB 395 ACAGTCAAGGTGCGCAGCGCAGATGTATCCACCGTGTGCTCCGCGCACCACCGCGCG 336
QY 610 GGCCAAAGGATCGCCCGCTGAGGCTTTACGAAGCCCGCGGAGTACGCGCGGGTTCCGA 669
DB 335 GGCCAAAGGATCGCCCGCTGAGGCTTTACGAAGCCCGCGGAGTACGCGCGGGTTCCGA 276
QY 670 TGATGTCCGCTGTGCGAGCCCGAGAGTGTGAGCATCTTCTCCGCGCGGACCACTACT 729
DB 275 TGATGTCCGCTGTGCGAGCCCGAGAGTGTGAGCATCTTCTCCGCGCGGACCACTACT 216
QY 730 AGCTACCATGTTAAAGCGAGTCCGCGCGAGTGTCAAGACGACGATGTGTACTGTGCGC 789
DB 215 AGCTACCATGTTAAAGCGAGTCCGCGCGAGTGTCAAGACGACGATGTGTACTGTGCGC 156
QY 790 GTCAAAATCCAGAAATGACGTAGTCTTGAATGGGCTCGCAATATGTGCGGTTCTGTTA 849
DB 155 GTCAAAATCCAGAAATGACGTAGTCTTGAATGGGCTCGCAATATGTGCGGTTCTGTTA 96

FEATURES
source
  Location/Qualifiers
    1..900
      /organism="Zea mays"
      /mol_type="mRNA"
      /db_xref="MaizeDB:635953"
      /db_xref="taxon:4577"
      /clone_lib="Maize Mapping Project/DuPont Cornsensus
      Library"
      /note="this sequence is part of a project of EST
      assemblies resulting from the application of public
      contigs to seed DuPont contigs; this resource was
      assembled by Dupont as part of a collaboration for the
      overgo addressing of BACs in conjunction with the Maize
      Mapping Project"
BASE COUNT    191 a    244 c    331 g    134 t
ORIGIN
Query Match      94.0%; Score 893; DB 11; Length 900;
Best Local Similarity 100.0%; Pred. No. 1.9e-80;
Matches 893; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 AGCATCAACAACAGTGGCGACAGACAAGATTGTGAGGGTGATCCGCGCTGAGAAGAG 110
DB 8 AGCATCAACAACAGTGGCGACAGACAAGATTGTGAGGGTGATCCGCGCTGAGAAGAG 67
QY 111 ATGGCTTAAGATCCGCGCGCGCGCGCGCGCTGTGCTTCGCGGCCCTGTGGCGCGTG 170
DB 68 ATGGCTTAAGATCCGCGCGCGCGCGCGCGCTGTGCTTCGCGGCCCTGTGGCGCGTG 127
QY 171 GCGGCTCTCCAGGCGAGGTCTGAGCGCGAGAGGCTCAGGACCTGAGTGTCTGGCAGGAG 230
DB 128 GCGGCTCTGCAAGGCGAGGTCTGAGCGCGAGAGGCTCAGGACCTGAGTGTCTGGCAGGAG 187
QY 231 GTCCAGGAGAGCCGCTCGACCGCTGCGCCAGGTCCTCGACCGGAGCTAACCGCGCGC 290
```

Db 188 GTCCAGGAGAGCCCGTTCGACGCGTCCGCGAGTCTCTGACCGGAGCTAAACCGCGGC 247
Qy 291 GCGCGCGCGCGCGCGTTCGCGCGTTCGCGTGGGACCGGGCTCCGGATCGGTGCTGC 350
Db 248 GCGCGCGCGCGCGCGTTCGCGCGTTCGCGTGGGACCGGGCTCCGGATCGGTGCTGC 307
Qy 351 CAGCAGCTCCAGGACGTGAGCGCGAGTCCCGTGGCGCGCCATCCGAGCATGTGTCAGG 410
Db 308 CAGCAGCTCCAGGACGTGAGCGCGAGTCCCGTGGCGCGCCATCCGAGCATGTGTCAGG 367
Qy 411 GCGTACGAGGCGCATGCGCGCGTTCGAGAAAGGCTGTGGCCATGCGGGCGGCGAGG 470
Db 368 GGTACGAGGAGCCATGCGCGCGTTCGAGAAAGGCTGTGGCCATGCGGGCGGCGAGG 427
Qy 471 CAGCGCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 530
Db 428 CAGCGCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 487
Qy 531 CCAGGAGAGGATACGGCTACGGTACAGGCTGCGCGAGCGGAGATGATCCACCGTGTCT 590
Db 488 CCAGGAGAGGATACGGCTACGGTACAGGCTGCGCGAGCGGAGATGATCCACCGTGTCT 547
Qy 591 CCGGACACCGCGCGCGCGGAGGATCGCGCGGCTGAGGCTTACGAAGCGCGCGGAG 650
Db 548 CCGGACACCGCGCGCGCGGAGGATCGCGCGGCTGAGGCTTACGAAGCGCGCGGAG 607
Qy 651 TAGCGCGCGGGTTCGCGATGATGTCGCGCTGTCGAGCGCCAGAGGTCAGCATCTTC 710
Db 608 TAGCGCGCGGGTTCGCGATGATGTCGCGCTGTCGAGCGCCAGAGGTCAGCATCTTC 667
Qy 711 TCCGCGCGCGACAGTACCTAGCTTACCATGTTAAAGCGAGTCCGCGCGAGGTCGAAGCG 770
Db 668 TCCGCGCGCGACAGTACCTAGCTTACCATGTTAAAGCGAGTCCGCGCGAGGTCGAAGCG 727
Qy 771 CAGCATGTGATCTGTCGCGTCAAAATCCAGATGACGTAGCTTGAAGTGGGCTCGCAA 830
Db 728 CAGCATGTGATCTGTCGCGTCAAAATCCAGATGACGTAGCTTGAAGTGGGCTCGCAA 787
Qy 831 TATTGTCGCGTTCGTTACATAATGATAATGATAATGATAATGATAATGATAATGATAATG 890
Db 788 TATTGTCGCGTTCGTTACATAATGATAATGATAATGATAATGATAATGATAATGATAATG 847
Qy 891 GCCAGATGATCTGCGCGCGTTCCTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 943
Db 848 GCCAGATGATCTGCGCGCGTTCCTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 900

RESULT 3
CC349105
LOCUS
DEFINITION
CC349105
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1038)
White, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick
, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek
, R.W., Nunberg, A., Robbins, D., and Lakey, N.
Consortium for Maize Genomics
Unpublished
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org

Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..1038
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMB0422E04"
/clone_lib="ZM 0.7 1.5 KB"
/note="Vector: pBSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"
BASE COUNT 215 a 316 c 336 g 171 t
ORIGIN
Query Match 77.1%; Score 732; DB 29; Length 1038;
Best Local Similarity 99.7%; Pred. No. 1.7e-64;
Matches 754; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
Qy 1 AAAAAAAAACCCCCTCGTCGATCACCAACAAAGAACACAGTAACCTAGCAGCTAGCACATCAA 60
Db 283 AAAAAAAAACCCCCTCGTCGATCACCAACAAAGAACACAGTAACCTAGCAGCTAGCACATCAA 342
Qy 61 ACAAGTGGCGACAGACAAAGATTGTGAGGGTGATCCGCGCTGAGAAAGATGGCTAAGA 120
Db 343 ACAAGTGGCGACAGACAAAGATTGTGAGGGTGATCCGCGCTGAGAAAGATGGCTAAGA 402
Qy 121 TCGCCGCGCGCGCGCGCGCGCTGTCTTCGCGCGCGCTGTGCGCGCGCTGTGCGCGCTGTGCC 180
Db 403 TCGCCGCGCGCGCGCGCGCGCTGTCTTCGCGCGCGCTGTGCGCGCGCTGTGCGCGCTGTGCC 462
Qy 181 AAGGCGAGGTCCAGCGCGCAGAGCGCTCAGGAGCTTCAGTGTCTGGCAGGAGGTCCAGGAGA 240
Db 463 AAGGCGAGGTCCAGCGCGCAGAGCGCTCAGGAGCTTCAGTGTCTGGCAGGAGGTCCAGGAGA 522
Qy 241 GCCCGCTCGACCGCTCCGCGCGCAGGTCTTCGACCGGACGCTAACCGCGCGCGCGCGCGCG 300
Db 523 GCCCGCTCGACCGCTCCGCGCGCAGGTCTTCGACCGGACGCTAACCGCGCGCGCGCGCGCG 582
Qy 301 GCGCGGTTGGCGCGCTTCGCGTGGCGCACCGGCTCCGATGCGGTGCTGCCAGCAGCTCC 360
Db 583 GCGCGGTTGGCGCGCTTCGCGTGGCGCACCGGCTCCGATGCGGTGCTGCCAGCAGCTCC 642
Qy 361 AGGAGTGGCGCGCGAGTCCGCTCGCGCGCCATCCGAGCATGTTCAGGGGTACGAGG 420
Db 643 AGGAGTGGCGCGCGAGTCCGCTCGCGCGCCATCCGAGCATGTTCAGGGGTACGAGG 702
Qy 421 AGGAGTGGCGCGCGAGTGGTGGCCATGGGCGCGCGCAGCAGCAGCGCGCGCG 480
Db 703 AGGAGTGGCGCGCGCTGGAGAAAGGCTGGTGGCCATGGGCGCGCGCAGCAGCAGCGCGCG 762
Qy 481 CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540
Db 763 CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 822
Qy 541 GATACGGCTACCGTCCAGGCTGCGCGCGCAGGCGCAGATGTATCCACCGTGTCTGCCGCGACCA 600
Db 823 GATACGGCTACCGTCCAGGCTGCGCGCGCAGGCGCAGATGTATCCACCGTGTCTGCCGCGACCA 882
Qy 601 CCGCGCGCGCGCGCAAGGATCCGCGCGCTGAGGCTTACGAAGGCCCGCGCGCGCGCGCG 659
Db 883 CCGCGCGCGCGCGCAAGGATCCGCGCGCTGAGGCTTACGAAGGCCCGCGCGCGCGCGCGCG 942
Qy 660 GGGTTGCCGATGATGTGCGCGCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 718
Db 943 GGGTTGCCGATGATGTGCGCGCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1002
Qy 719 CGACGAGTACTAGCTACCATGCTTAAAGCGAGTCCG 754
Db 1003 CGACGAGTACTAGCTACCATGCTTAAAGCGAGTCCG 1038
RESULT 4

Db 322 AGGAGAGGATACGGCTACGGTACAGAGTGGCCAGCAGCAGATGATCCACCGTGTGGTCC 381
 Qy 593 CGGCACACCGCGGGGGCCCAAGGATCGCCCGGTAGGCTTACGAAGCCCGGAGTA 652
 Db 382 CGGCAC-----CGGGCAAAATCGCCCGTGAAGCTTACAAAGGCCAGGAGTA 432
 Qy 653 CGCGCGGGGTTCGGATGATGTCGGCTGTCGGAGCCCGCAGAGTGCAGCATCTTCTC 712
 Db 433 CGCGCAGGATCCGATGATGTCGGCTGTCGGAGCCCGCAGAGTGCAGCATCTTCTC 492
 Qy 713 CGCGCGGACACGATGATGTCGGCTGTCGGAGCCCGCAGAGTGCAGCATCTTCTC 772
 Db 493 CGCGCGGACAGTA---CTACTAGGTAAAGTGAAGTGGCGCGCCGTGCAAGCGCA 548

RESULT 11
 BE917909
 LOCUS
 DEFINITION
 BE917909
 VERSION
 BE917909.1 GI:10420483
 KEYWORDS
 SOURCE
 ORGANISM
 Sorghum bicolor (sorghum)
 Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoidae; Andropogoneae; Sorghum.
 REFERENCE
 AUTHORS
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 L.H.
 TITLE
 JOURNAL
 COMMENT
 An EST database from Sorghum: ovaries of varying immature stages
 Unpublished
 Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 523
 POLYA=No.

FEATURES
 source
 1. .605
 /organism="Sorghum bicolor"
 /mol_type="mRNA"
 /db_xref="taxon:4558"
 /clone_lib="Ovary 1 (OV1)"
 /notes="Organ: Mix of ovaries of varying immature stages
 from 8-week-old plants; Vector: pBluescript II from lambda
 Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
 from poly-A RNA in the cloning vector lambda Zap II.
 Clones to be sequenced were prepared by mass excision."
 122 a 184 c 217 g 82 t

BASE COUNT
 ORIGIN
 Query Match 38.2%; Score 362.6; DB 10; Length 605;
 Best Local Similarity 80.9%; Pred. No. 1.2e-27;
 Matches 520; Conservative 0; Mismatches 64; Indels 59; Gaps 6;
 Qy 40 AACTAGCAGTACACATCAAAAGTGGCGCAGACAAAGATTTGTGAGGTGATCCGC 99
 Db 19 AACAATAGTACAGATCAAACTAGCAGACAAAGATTTGTATGA-----TCC 70
 Qy 100 GCTGAGAAGAGTGGCTAGATC---GCCGCGCGCGCGCGCGCTGTGTTCCGCG 156
 Db 71 GTGCTGATCAGATGGCTAAGATCGCTCGCGTCCGCCACCGCGCGCTGTGCTCGCG 130

Qy 157 CCTGTGTCGGTGGCGCTCTGCCAAGCGAGGTCCAGCGCAGAGGCTCAGGACCTGC 216
 Db 131 CCTGTGTCGGTGGCGCTCTGCCAAGCGAGGTCCAGCGCAGAGGCTCAGGACCTGC 190
 Qy 217 AGTGTGTCAGGAGGTTCAGGAGAGCCCGCTCGACCGCTGCCCGCAGGCTCTCGACCGC 276
 Db 191 AGTGTGTCAGGAGGTTCAGGAGAGCCCGCTCGCGCGCTGCCCGCAGGCTCTCGACCGC 250
 Qy 277 AGCTAACCGCGCGCGCGCGCGCGCGCTTGGCCCGCTTCCGTTGGGCGACCGGGCTCC 336
 Db 251 AGCTCAC---CGCGCGCATGCGCTACGGCATCGGCCCGCTTCCGTTGGGCGACCGGGCTCC 307
 Qy 337 CGATGCGGTGTCGCCAGCAGCTCCAGGAGCTGAGCGCGAGTCCCGCTCGCGCGCATCC 396
 Db 308 GATGCGGTGTCGCCAGCAGCTCCAGGAGCTGAGCGCGAGTCCCGCTCGCGCGCATCC 367
 Qy 397 GGAGCATGCTCAGGGGCTACGAGAGGCGCATCCCGCGCTGGAGAAAGCTGTGGCCAT 456
 Db 368 GGAGCATGCTCAGGGGCTACGAGAGGCGCATCCCGCGCTGGAGAAAGCTGTG-----T 421
 Qy 457 GGGGGCGCGCAGCAGCAGCGCGCGCGCGCGCGAGGAGGAGGAGGACAGGGGGCTACTACT 516
 Db 422 GGGGGCAGCAGCG 451
 Qy 517 ACCCTGTCAGCG 576
 Db 452 ACCCTGTCAGCG 511
 Qy 577 ATCCACCGTGTCTCCCGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 636
 Db 512 ACCACCGTGTCTCCCGCAC-----CGGGCAAAATCGCGCGCGCGCGCGCGCGCGCG 562
 Qy 637 CGAAGCG 679
 Db 563 CAAAGCG 605

RESULT 12
 BM500544
 LOCUS
 DEFINITION
 PAC00000000895 Pioneer AF-1 array Zea mays cDNA, mRNA sequence.
 BM500544
 VERSION
 BM500544.1 GI:18659831
 KEYWORDS
 EST.
 SOURCE
 Zea mays
 Zea mays
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoidae; Andropogoneae; Zea.
 REFERENCE
 AUTHORS
 Hunter,B.G., Beatty,M., Singletary,G., Hamaker,B., Larkins,B.A. and
 Jung,R.
 TITLE
 Maize opaque endosperm mutations create extensive changes in
 patterns of gene expression
 JOURNAL
 Unpublished
 COMMENT
 Contact: Jung R
 Trait and Technology Development, Food and Feed Research
 Pioneer Hi-Bred International, Inc.
 7300 NW 62nd Ave., P.O. Box 1004, Johnston, IA 50131-1004, USA
 Tel: 515 270 5934
 Fax: 515 254 2619
 Email: rudolf.jung@pioneer.com.
 FEATURES
 source
 1. .367
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="taxon:4577"
 /clone_lib="Pioneer AF-1 array"
 BASE COUNT 72 a 109 c 130 g 51 t 5 others
 ORIGIN
 Query Match 35.8%; Score 340.2; DB 12; Length 367;
 Best Local Similarity 97.8%; Pred. No. 2.9e-25;

Db 275 AAAAAAAAAACCTCATCGATCACCAACCAAGAACACAGTAACCTAGCTAGCTAGCAGATCAA 334
Qy 61 ACAAGTGGGACAG-----ACAAGATTGTGAGGGTGATCC-CGGCTGAGAAGAGATGGC 115
Db 335 ACAAGTGGGACAGCTCTACAAGATTGTGAGAGTGATCCGGCGCTGAGAAGAGATGGC 394
Qy 116 TAAGATCGCCCGCGCGCGCGCGC---GGCGCTGTGCTTCCGGCGCCTGTGTGCGCGTGGC 172
Db 395 TAAGATCGCCCGCGCGCGCGCGCGCTGTGCTTCCGGCGCCTGTGTGCGCGTGGC 454
Qy 173 CGTCTGCCAAGGCGAGGTGAGCGCGCAGAGCGCTCAGGACCTCAGTGTGCGCAGAGGT 232
Db 455 CGTCTGCCAAGGCGAGGTGAGCGCGCAGAGCGCTCAGGACCTCAGTGTGCGCAGAGGT 514
Qy 233 CAGGAGAGCGCGCTGACCGCTGCGCGCAGAGTCTCGACCGCAGCTAACCGCGCGCG 292
Db 515 CAGGAGAGCGCGCTGACCGCTGCGCGCAGAGTCTCGACCGCAGCTAAC-----565
Qy 293 CGCGCGCGCGCGCTTGGCGCGCTTCCGGTGGCGCACCGCGCTCCGGATGCGGTGCTGCCA 352
Db 566 CGGTGGCGCGCGCTGCGCGCGCTTCCGGCGCGCACCGCGCTCCGGATGCGGTGCTGACA 625
Qy 353 GCAGCTCCAGGACGTGAGCGCGAGTG 379
Db 626 GCAGCTCCAGGACGTGAGCGCGAGTG 652

RESULT 15

BG048804
LOCUS OVI_23_C05.b1_A002 Ovary 1 (OVI) Sorghum bicolor cDNA, mRNA
DEFINITION 425 bp mRNA linear EST 25-JAN-2001
sequence.
ACCESSION BG048804
VERSION BG048804.1 GI:12499929
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 425)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.

Seq primer: JEN REV

High quality sequence stop: 393

POLYA=No.

FEATURES

source

Location/Qualifiers

1..425
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Ovary 1 (OVI)"
/notes="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from Lambda
Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."
89 a 127 c 149 g 60 t

BASE COUNT
ORIGIN

Query Match

27.0%; Score 256.2; DB 10; Length 425;

Best Local Similarity 82.5%; Pred. No. 5.8e-17;
Matches 362; Conservative 0; Mismatches 28; Indels 49; Gaps 4;
Qy 334 TCCGATGCGGTGCTGCCAGCAGCTCCAGGACGTGAGCCGCGAGTCCCTGCGCCGCCA 393
Db 1 TCCGATGCGGTGCTGCCAGCAGCTCCAGGACGTGAGCCGCGAGTCCCTGCGCCGCCA 60
Qy 394 TCCGAGCATGCTCAGGGGTACGAGGAGCCATCCCGCGCTGGAGAAAGCTGTGTGC 453
Db 61 TCCGAGCATGCTCAGGGGTACGAGGAGCCATCCCGCGCTGGAGAAAGCTGTGTGC 116
Qy 454 CATGGGGCGCGCAGCAGCAGCCCGCCCGCAGGAGGAGGACAGGAGGAGGCTACT 513
Db 117 --TGGGGGCGAGCCGCGAGCC-----GGGCTAGC 144
Qy 514 ACTACCCCTGCGAGCGCGCCAGGAGGAGGATACGGCTACGGTCAAGGTGCGCCAGCGCAGA 573
Db 145 ACTACCCCTGCGAGCCAGGAGGAGGATACGGCTACGGTCAAGGTGCGCCAGCAGAGA 204
Qy 574 TGTATCCACCGTGTGTCCTCCGGCACCCCGCGCGGCGCAAGGATCCGCGCGTGAGGC 633
Db 205 TGTATCCACCGTGTGTCCTCCGGCAC-----CGGGCAAAAAATCGCCCGGTGAAGC 255
Qy 634 TTACGAAGCCCGCGAGTACGCGCGGGGTTCCTCATGATGTCCGGCTGTTCGAGCCCC 693
Db 256 TTACGAAGCCCGCGAGTACGCGCGGGATGCCGATGATGTCCGGCTGTTCGAGCCCC 315
Qy 694 AGGAGTGCAGCATCTTCTCCCGCGCGCAGCAGTACTAGTACCATGTTAAAGCGAGTCG 753
Db 316 AGGAGTGCAGCGCTCTTCTCCCGCGCGCAGCAGTA----CTACTAGGGTAAGAGTGTAGTCG 371
Qy 754 GCGCGAGGTGCAAGAGCGCA 772
Db 372 GCGCGCGGTGCAAGAGCGCA 390

Search completed: November 29, 2003, 15:52:57

Job time : 2422 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 29, 2003, 17:07:40 ; Search time 3510 Seconds
(without alignments)
2400.964 Million cell updates/sec

Title: US-10-053-410-4
Perfect score: 1148
Sequence: 1 MAKIAAAAAALCPAALVAV.....MMCRILSEPQECIFSGGDQY 206

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh

-Q=/cgn2_1/USFTO.spool/US10053410/runat_28112003.140756_27880/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=us10053410 @CGN 1.1 3508 @runat_28112003.140756_27880 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.or.*
21: em.ro.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.nam.*
37: em.htg.vrt.*
38: em.sv.*
39: em.hgo.hum.*
40: em.hgo.mus.*
41: em.hgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1148	100.0	932	8	AF371278 Zea mays
2	439	38.2	1773	8	D50643 Oryza sativ
3	439	38.2	99538	2	AC113332 Oryza sat
4	439	38.2	124132	2	AC130605 Oryza sat
5	435	37.9	757	8	X6390 O.sativa mr
6	433.5	37.8	825	8	LI252 Oryza sativ
7	414	36.1	102842	8	AF497474 Aegilops
8	381	33.2	120562	8	AY268139 Hordeum v
9	292.5	25.5	1962	8	AJ306974 Aegilops
10	289.5	25.2	3095	8	X03041 Wheat gene
11	289	25.2	2163	8	AY245797 Triticum
12	289	25.2	2330	8	AJ34785 Triticum
13	287	25.0	1980	8	AY263343 Thinopyru
14	287	25.0	2996	8	X61026 Wheat Glu-1
15	286	24.9	906	8	AY263345 Thinopyru
16	286	24.9	2220	8	AY263346 Thinopyru
17	285	24.8	1494	8	AY299518 Thinopyru
18	283	24.7	2409	8	AF513640 Aegilops
19	282.5	24.6	1859	8	HVDNAHOR3
20	282	24.6	1494	8	AY264065 Thinopyru
21	282	24.6	1788	8	AY298724 Thinopyru
22	282	24.6	2809	8	ASU39229 Thinopyru
23	281.5	24.5	2296	8	BLYHOR3
24	281.5	24.5	2380	6	BD012675 Method of
25	281.5	24.5	2434	6	BD012674 Method of
26	281	24.5	6462	8	TAGLD12B
27	279	24.3	2804	8	AY248704 Aegilops
28	277.5	24.2	598	8	TAE308967 Triticum
29	276.5	24.1	1371	8	AF476962 Aegilops
30	275	24.0	2331	8	SCB314781 Secale ce
31	273.5	23.8	1881	8	AY174159 Aegilops
32	271	23.6	1842	8	AY263344 Thinopyru
33	269	23.4	2297	8	AY249141 Triticum
34	267.5	23.3	2403	8	SCB314770 Secale ce
35	267	23.3	1920	8	AY245579 Triticum
36	267	23.3	1998	8	ACY306973 Aegilops
37	267	23.3	2313	8	SCB314767 Secale ce
38	267	23.3	2331	8	AF216869 Triticum
39	267	23.3	2331	8	SCB314775 Secale ce
40	267	23.3	2331	8	SCB314777 Secale ce
41	267	23.3	2331	8	TAE314783 Triticum
42	264.5	23.0	537	8	AF226698 Aegilops
43	262.5	22.9	1905	8	AF354289 Aegilops
44	260	22.6	1905	8	AF476960 Aegilops
45	259.5	22.6	1830	8	AY245578 Triticum

ALIGNMENTS

RESULT 1

AF371278	AF371278	932 bp	mRNA	linear	PLN 22-OCT-2001
LOCUS	Zea mays alpha globulin mRNA, complete cds.				
DEFINITION	Zea mays alpha globulin mRNA, complete cds.				
ACCESSION	AF371278				
VERSION	AF371278.1	GI:16305141			
KEYWORDS					
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.				
AUTHORS	Woo, Y.M., Hu, D.W., Larkins, B.A. and Jung, R.				
TITLE	Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression				
JOURNAL	Plant Cell 13 (10), 2297-2317 (2001)				
MEDLINE	2148069				
PUBMED	11595803				
REFERENCE	2 (bases 1 to 932)				
AUTHORS	Jung, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-APR-2001) TTD, Pioneer Hi-Bred International, Inc., 7300NW 52nd Ave., Johnston, IA 50131, USA				
FEATURES	Location/Qualifiers				
source	1. .932				
	/organism="Zea mays"				
	/mol_type="mRNA"				
	/variant="B73"				
	/db_xref="taxon:4577"				
	/tissue_type="endosperm"				
	108. .728				
	/codon_start=1				
	/product="alpha globulin"				
	/protein_id="AAL16993.1"				
	/db_xref="GI:16305142"				
	/translation="MAKATAAAALCFALVAVAVCGEVEORLRLDLCQWQEVES PLDARQLVDRLQITGGGGGGVPPRWGTGMRCCQLOQDVRECAAIRSMVRGY EAMPPLERKQWPRQGGGGGGYPPCRSPGEGYGGGGGQGGQRYPPCR PGTTGGPRIGRVLTKAREYAAAGLPMWCRSLSEFQCSIFSGDQY"				
CDS					
	BASE COUNT 199 a 259 c 334 g 140 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	1.13e-51	Length:	932		
Score:	1148.00	Matches:	206		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	8	Gaps:	0		
US-10-053-410-4 (1-206) x AF371278 (1-932)					
QY	1 MetAlaIyIleAlaAlaAlaAlaAlaAlaLeuCySPheAlaAlaLeuValAlaVal 20				
Db	108 ATGCTTAAGATCCGCGCGCGCGCGCGCTGTCTTCGCGCGCTGTGTGCGCGTG 167				
QY	21 AlAlaIyCysAlnGlyValGluAlaGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40				
Db	168 GCGCTGTCAGAGCGAGTGTGAGCGCGAGAGCTCAGGACCTCGAGTGTGCGAGG 227				
QY	41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly 60				
Db	228 GTCCAGAGAGCCGCTCGACGCGTCCGCGCAGAGTCTCGACCGCGAGCTAACCGCGGC 287				
QY	61 GlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys 80				
Db	288 GCGCGCGCGCGCGCGCTTGGCCGCTTCCGCTGCGGCGACCGCGCTCCGAGTGTGTCG 347				
QY	81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100				
Db	348 CAGCAGCTTCAGAGCTGAGCGCGAGAGTCCGCTCGCGCGCATCCGAGAGCATGTGTCAGG 407				
QY	101 GlyTyTrpGluAlaMetProLeuGluGlyTyTrpTrpTrpTrpTrpTrpTrpTrpTrp 120				

```
LDACROVLDRQLTGRFRFOPFRFPALGLRMCCQOQLQDVSRRCRAAIRMRVSYE
EMPMLEGGWSSSEYQYGGSEGGYGGSEGGYGGQOQQPQWTRVRLTRA
ROYAALPQMCVRPEQCSIFAAGQY"
BASE COUNT 471 a 448 c 448 g 406 t
ORIGIN

Alignment Scores:
Pred. No.: 1,03e-14 Length: 1773
Score: 439.00 Matches: 111
Percent Similarity: 61.61% Conservative: 19
Best Local Similarity: 52.61% Mismatches: 52
Query Match: 38.24% Indels: 30
DB: 8 Gaps: 9

US-10-053-410-4 (1-206) x D50643 (1-1773)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
Db ATGGCTAGCAAGTCTCTTCTTCGGCGGCGCTC---ATGGCGGCCATGGTGGCCATC 1065
QY 21 AlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40
Db TCCGGCGCGCAGTGCAGCGAGTCCGAGATCAGGTTTCAGGACAGGAGTCCAGCGGGAG 1125
QY 41 ValGlnGluSerProLeuAspAlaCysAsgGlnValLeuAspArgGlnLeuThrGly-Gl 60
Db GTGCAGGACAGCCCGCTGCAGCGTCCGCGAGTCTCGACCGGAGTCTACCGCGCGG 1185
QY 60 YGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCy 80
Db GAGAGTTCACGCGCGAGTTCGCCCGCCCGCGGCG-CGCTCGGCGCTCGGATGCGTGTG 1244
QY 80 sGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValAr 100
Db CCAGCAGCTGCAGGAGTGCAGCGCGAGTCCGCTCGCGCGCATCCGCGGATGTGTAG 1304
QY 100 gGlyTrpGluGluAlaMetPro---ProLeuGluLysGlyTrpTrpProTrpGlyArgGl 119
Db GAGCTACGAGGAGAGCATCCGATGCCCTTGAGCAAGGCTGCTCGTCTGTCGTCGCGA 1364
QY 119 nGlnGlnProProGlnGlyGlyGlyGlyGlnGlyGlnGlyGlyTyTrpTyTrpProCysSe 139
Db GTACTATC-----GCGCGCGAGGGG-----TC 1385
QY 139 rArgProGlyGluGlyTyTrpGlyGlyGlyGlyGlnArgGlnMetTyr-----156
Db GTCTCGGAGAGGGGTAC---TACGCGAGGGGTCTCGAGAGGGGTACTACGCGCA 1442
QY 157 -ProProCysArgProGlyThrThrGlyGlyGlyProArgIleGlyArgValArgLeuTh 176
Db GCAGCAGCAGCAGCGCGGGATGACC-----CGCTGAGGCTGAC 1481
QY 176 rIysAlaArgGluTyTrpAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGl 196
Db CAGGCGCGCAGCAGTACGCGCGCAGTCCGCTCGATGTGCGGGTT---GAGCCCCAGCA 1538
QY 196 uCysSerIlePheSerGlyGlyAspGlnTyr 206
Db GTGCAGCATCTTCGCCCGCGC---CAGTAC 1566

RESULT 3
LOCUS AC113332/c
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 5 clone
CUI057_B02, ** SEQUENCING IN PROGRESS ***, 5 ordered pieces.
ACCESSION AC113332
VERSION AC113332.1 GI:19033375
KEYWORDS HTG; HTGS; PHASE2.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
BASE COUNT 26954 a 22452 c 22701 g 27030 t 401 others
ORIGIN
Alignment Scores:
Pred. No.: 3,05e-13 Length: 99538
Score: 439.00 Matches: 111
Percent Similarity: 61.61% Conservative: 19
Best Local Similarity: 52.61% Mismatches: 52
Query Match: 38.24% Indels: 30
DB: 2 Gaps: 9

US-10-053-410-4 (1-206) x AC113332 (1-99538)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
Db ATGGCTAGCAAGTCTCTTCTTCGGCGGCGCTC---ATGGCGGCCATGGTGGCCATC 34052
QY 21 AlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40
Db TCCGGCGCGCAGTGCAGCGAGTCCGAGATCAGGTTTCAGGACAGGAGTCCAGCGGGAG 33992
QY 41 ValGlnGluSerProLeuAspAlaCysAsgGlnValLeuAspArgGlnLeuThrGly-Gl 60
Db GTGCAGGACAGCCCGCTGCAGCGTCCGCGAGTCTCGACCGGAGTCTACCGCGCGG 33932
QY 60 YGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCy 80
```

```
1 (bases 1 to 99538)
Chow, T.-Y., Hsiao, Y.-I. C., Chen, C.-S., Chen, H.-H., Wu, H.-P.,
Liu, S.-M., Chao, Y.-T., Chang, S.-J., Chen, T.-R., Chen, Y.-L.,
Chow, M.-H. J., Hong, Y.-C., Hsiung, J.-N., Hsu, C.-H., Huang, J.-J.,
Kau, P.-I., Lee, M.-C., Leu, H.-L., Lin, S.-J., Wu, L.-F. and Shaw, J.-P.
Oryza sativa BAC OJ1057_B02 genomic sequence
Unpublished
2 (bases 1 to 99538)
Chow, T.-Y. and Hsiao, Y.-I. C.
Direct Submission
Submitted (01-MAR-2002) Institute of Botany, Academia Sinica, 128,
Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
3 (bases 1 to 99538)
Hsiao, Y.-I. C. and Chow, T.-Y.
Direct Submission
Submitted (29-MAR-2003) Institute of Botany, Academia Sinica, 128,
Section 2, Yen-chu-yuan Road, Nankang, Taipei 11529, Taiwan
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and ASPGC-Taiwan sequencing data.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 890: contig of 890 bp in length
891 990: gap of unknown length
991 19703: contig of 18713 bp in length
19704 19803: gap of unknown length
19804 75538: contig of 55735 bp in length
75539 75638: gap of unknown length
75639 92490: contig of 16852 bp in length
92491 92590: gap of unknown length
92591 99538: contig of 6948 bp in length.
Location/Qualifiers
1. 99538
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="5"
/clone="OJ1057_B02"
/notes="japonica cultivar-group"
BASE COUNT 26954 a 22452 c 22701 g 27030 t 401 others
ORIGIN
Alignment Scores:
Pred. No.: 3,05e-13 Length: 99538
Score: 439.00 Matches: 111
Percent Similarity: 61.61% Conservative: 19
Best Local Similarity: 52.61% Mismatches: 52
Query Match: 38.24% Indels: 30
DB: 2 Gaps: 9

US-10-053-410-4 (1-206) x AC113332 (1-99538)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
Db ATGGCTAGCAAGTCTCTTCTTCGGCGGCGCTC---ATGGCGGCCATGGTGGCCATC 34052
QY 21 AlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40
Db TCCGGCGCGCAGTGCAGCGAGTCCGAGATCAGGTTTCAGGACAGGAGTCCAGCGGGAG 33992
QY 41 ValGlnGluSerProLeuAspAlaCysAsgGlnValLeuAspArgGlnLeuThrGly-Gl 60
Db GTGCAGGACAGCCCGCTGCAGCGTCCGCGAGTCTCGACCGGAGTCTACCGCGCGG 33932
QY 60 YGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCy 80
```

Db 33931 GAGAGGTTCAGCGCATGTTCCCGCCGCCGCGGCGCTCGGATCGAGTGCCTG 33873

Qy 80 sGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaLeuValSerMetValar 100
|||||
|||||

Db 33872 CCAGCAGCTCAGGACGTGAGCGCGAGTCCGCTCGCGCCCATCGCGGATGTGTAG 33813
|||||
|||||

Qy 100 gGlyTyrGluGluAlaMetPro---ProLeuGluLysGlyTyrTrpProTyrGlyArgG1 119
|||||
|||||

Db 33812 GAGCTACGAGAGAGCATGCGATGCCCTCGGACAGGCTGTGCTGCTGCTGCTGCGGA 33753
|||||
|||||

Qy 119 nGlnGlnProProGlnGlnGlyGlyGlyGlnGlnGlyTyrTyrProCys 139
|||||
|||||

Db 33752 GTACTAC-----GCGCGGAGGGG-----TC 33732
|||||
|||||

Qy 139 rArgProGlyGluGlyTyrGlyTyrGlyGlnGlnGlyGlnArgGlnMetTyr----- 156
|||||
|||||

Db 33731 GTGCTCGGACAGGGGTAC---TACGCGGAGGGGTGCTCGGAGGAGGCTTACTACGCGGA 33675
|||||
|||||

Qy 157 -ProProCysArgProGlyThrThrGlyGlyGlyProArgIleGlyValArgValArgLeuTh 176
|||||
|||||

Db 33674 GCAGCAGCAGCAGCGGGGATGACC-----CGCGTGAAGCTGAC 33636
|||||
|||||

Qy 176 rLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnG1 196
|||||
|||||

Db 33635 CAGCGCGAGGAGTACGCGCGCAGCTGCGCTGATGTGCGGGTT---GAGCCCCAGCA 33579
|||||
|||||

Qy 196 uCysSerIlePheSerGlyAspGlnTyr 206
|||||
|||||

Db 33578 GTGCAGCATCTCGCGCGCGC---CAGTAC 33551
|||||
|||||

RESULT 4

AC130605/c 124132 bp DNA linear HTG 13-AUG-2002
LOCUS
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 5 clone P0010D04,
*** SEQUENCING IN PROGRESS ***, 4 ordered pieces.

AC130605
AC130605.1 GI:22212956
HTG; HTGS PHASE2.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartodeae; Oryzaceae; Oryza.

REFERENCE

1 (bases 1 to 124132)
Chow,T.-Y., Hsing,Y.-I.C., Chen,C.-S., Chen,H.-H., Liu,S.-M.,
Chao,Y.-T., Chang,S.-J., Chen,H.-C., Chen,S.-K., Chen,T.-R.,
Chen,Y.-L., Cheng,C.-H., Chung,C.-I., Han,S.-Y., Hsiao,S.-H.,
Hsiung,J.-N., Hsu,C.-H., Huang,J.-J., Kau,P.-I., Lee,M.-C.,
Leu,H.-L., Li,Y.-F., Lin,S.-J., Lin,Y.-C., Wu,S.-W., Yu,C.-Y.,
Yu,S.-W., Wu,H.-P. and Shaw,J.-F.
Oryza sativa PAC P0010D04 genomic sequence

JOURNAL

Unpublished
2 (bases 1 to 124132)
Chow,T.-Y. and Hsing,Y.-I.C.

AUTHORS

Direct Submission

TITLE

Submitted (13-AUG-2002) Institute of Botany, Academia Sinica, 128,
Section 2, Academia Road, Nankang, Taipei 11529, Taiwan

JOURNAL

* NOTE: This is a 'working draft' sequence. It currently
consists of 4 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
provided by the submitter.

COMMENT

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 24826: contig of 24826 bp in length
* 24827 24926: gap of unknown length
* 24927 39423: contig of 14497 bp in length
* 39424 39523: gap of unknown length
* 39524 89605: contig of 50082 bp in length
* 89606 89705: gap of unknown length
* 89706 124132: contig of 34427 bp in length.

FEATURES

source

1. 124132
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="5"
/clone="P0010D04"
BASE COUNT 34211 a 27001 c 26916 g 35690 t 314 others
ORIGIN

Alignment Scores:

Pred. No.: 3 68e-13 Length: 124132
Score: 439.00 Matches: 111
Percent Similarity: 61.61% Conservative: 19
Best Local Similarity: 52.61% Mismatches: 52
Query Match: 38.24% Indels: 30
DB: 2 Gaps: 9

US-10-053-410-4 (1-206) x AC130605 (1-124132)

Qy 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
|||||
|||||

Db 119786 ATGGCTAGCAAGTCTCTTCTTCGCGCGCGGCTC---ATGCGCGCCATGTGTGCGCATC 119730
|||||
|||||

Qy 21 AlaValCysGlnGlyGluValGlnArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40
|||||
|||||

Db 119729 TCCGCGCGCAGCTGAGCGAGTGGAGATGAGTTCAGGACAGGAGTGCACGCGGAG 119670
|||||
|||||

Qy 41 ValGlnGlnSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly-G1 60
|||||
|||||

Db 119669 GTGCAGGACAGCCCGTGGACGCGTCCGCGAGGTGCTCAGCGGAGTCCACCGCGCG 119610
|||||
|||||

Qy 60 yGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCy 80
|||||
|||||

Db 119609 GAGAGTTCAGCCGATGTTCCGCCCGCGGCGGCGCTCGGCTCGGATGCGAGTGCCTG 119551
|||||
|||||

Qy 80 sGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaLeuArgSerMetValar 100
|||||
|||||

Db 119550 CCAGCAGCTGCGAGGAGCTGAGCGCGAGTGCCTCGCGCGCATCCGCGGATGTGTAG 119491
|||||
|||||

Qy 100 gGlyTyrGluGluAlaMetPro---ProLeuGluLysGlyTyrTrpProTyrGlyArgG1 119
|||||
|||||

Db 119490 GAGCTACGAGGAGAGCATGCGCATGCCCTGGAGCAAGGCTGCTGCTGCTGCTGCGGA 119431
|||||
|||||

Qy 119 nGlnGlnProProGlnGlnGlyGlyGlyGlnGlnGlyTyrTyrProCys 139
|||||
|||||

Db 119430 GTACTAC-----GCGCGGAGGGG-----TC 119410
|||||
|||||

Qy 139 rArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArgGlnMetTyr----- 156
|||||
|||||

Db 119409 GTGCTCGGAGCAGGGGTAC---TACGCGAGGGGTGCTCGGAGGAGGCTTACTACGCGGA 119353
|||||
|||||

Qy 157 -ProProCysArgProGlyThrThrGlyGlyGlyProArgIleGlyValArgValArgLeuTh 176
|||||
|||||

Db 119352 GCAGCAGCAGCAGCGGGATGACC-----CGCGTGAAGCTGAC 119314
|||||
|||||

Qy 176 rLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnG1 196
|||||
|||||

Db 119313 CAGGCGGAGGAGTACGCGCGCAGTGCCTGCGATGTGCGGGTT---GAGCCCCAGCA 119257
|||||
|||||

Qy 196 uCysSerIlePheSerGlyAspGlnTyr 206
|||||
|||||

Db 119256 GTGCAGCATCTTCGCGCGCGGCGC---CAGTAC 119229
|||||
|||||

RESULT 5

OS19KDGLO

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

OS19KDGLO 757 bp mRNA linear PLN 09-FEB-1999
O. sativa mRNA for 19 kDa globulin.
X63990.76663
X63990.1 GI:20158
19 kDa globulin; alpha-globulin; cereal storage protein; globulin.
Oryza sativa (japonica cultivar-group)

ORGANISM	Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
Ehrhartioideae; Oryzeae; Oryza.	
REFERENCE	1 (bases 1 to 757)
AUTHORS	Shorrosh,B.S., Wen,L., Zen,K.C., Huang,J.K., Pan,J.S., Hermanson,M.A., Tanaka,K., Muchukrishnan,S. and Reek,G.R.
TITLE	A novel cereal storage protein: molecular genetics of the 19 kDa globulin of rice
JOURNAL	Plant Mol. Biol. 18 (1), 151-154 (1992)
MEDLINE	92119226
PUBMED	1731968
REFERENCE	2 (bases 1 to 757)
AUTHORS	Shorrosh,B.
TITLE	Direct Submission
JOURNAL	Submitted (17-MAY-1992) B. Shorrosh, The Samuel Roberts Noble Foundation, plant Biology Division, 2510-1990East, Ardmore, Oklahoma 73402, USA
FEATURES	source 1..757 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="mRNA" /strain="Nipponbare (Japonica)" /db_xref="taxon:39947" /tissue_type="Endosperm" /dev_stages="Endosperm, immature" 29..589 CDS /product="19 kDa globulin precursor" /protein_id="CAA5400.1" /db_xref="GI:20159" /translat="WASKVVFPAALMAAMVAISGAHVSESMRFRDQCQREVODSPY LACQVLDRQLTGREFOPFRPGLRMQCCLODVSRRCRAAIRMRVRVE ESMPLEQSSESSEYEGSGSSSQGYGESSBEGYGGQQQPQGWTRVRLTRA RYAARLPFMSKRVFPQCSIFAAGQI"
sig_peptide	29..94
mat_peptide	95..586
polya_signal	/product="19 kDa globulin"
terminator	666..672
polya_signal	673..679
terminator	681..688
polya_signal	707..712
terminator	721..728
polya_site	742
BASE_COUNT	157 a 197 c 263 g 140 t
ORIGIN	
Alignment Scores:	
Pred. No.:	8.12e-15 Length: 757
Score:	435.00 Matches: 110
Percent Similarity:	61.1% Conservative: 19
Best Local Similarity:	52.1% Mismatches: 53
Query Match:	Indels: 30
DB:	Gaps: 9
US-10-053-410-4 (1-206) x OS19KDGL0 (1-757)	
Qy	1 MetAlaLyseIlleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
Db	29 ATGGCTAGCAGGTGCTCTTCTTCGCCGCGCGCTC---ATGGCGGCATGTTGCCCATC 85
Qy	21 AlaValCysGlnGlyGluValGluArgGlnArgLeuAspLeuGlnCysTrpGlnGlu 40
Db	86 TCCGGCGGCACGTGAGCGAGTCGAGATGAGTTTCAGGGACAGCGAGTCCAGCGGGAG 145
Qy	41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly-Gl 60
Db	146 GTGCAGGACAGCCCCTGGACCGTCCGCGCAGGTGCTCGACCGCGCACTCACCGCGCG 205
Qy	60 YGlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCy 80
Db	206 GAGAGGTTCCACCGCATGTTCCGCCCGCCCGGCG-CGCGTCCGCTCGCGGATGATGCTG 264

QY	80	sGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValAr	100
Db	265	CCAGCAGCTGCAGGACGTGAGCGCGAGTGCCTGCGCGCCATCCGCGGATGGTAG	324
QY	100	gGlyTyrlrGluaLaMetPro---ProLeuGluLysGlyTrpTrpProTrpGlyArgGl	119
Db	325	GAGCTACAGGAGAGCATGCCGATGCCCTGGACNAAGCTGGTCGTCGTCGTCGCGGA	384
QY	119	nGlnGlnProProGlnGlyGlyGlyGlyGlnGlyGlyTyrtyrtyrtyrtyrtyrtyr	139
Db	385	GTACTAC-----GGCGGCGAGGG-----TC	405
QY	139	tArgProGlyGluGlyTyrlrGlyGlyGlnGlyGlnArgGlnMetTyr---	156
Db	406	GTCTGCGGAGCAGGGGTAC---TACGGCGAGGGTCTCGGAGGAGGGCTACTACGGCG	462
QY	157	-ProProCysArgProGlyThrThrGlyGlyGlyProArgIleGlyArgValArgLeuth	176
Db	463	GCAGCAGCAGCAGCGGGGATGACC-----CGCGTGAGGCTGAC	501
QY	176	rlysAlaArgGluTyrlrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGl	196
Db	502	CAGGCGAGGACAGTACCGCGCGCAGCTGCCGTTCATGTCCGGGTT---GAGCCCCAGCA	558
QY	196	uCysSerIlePheSerGlyGlyAspGlnTyr 206	
Db	559	GTGCAGCATCTTCGCCGCGGC---CAGTAC 586	
RESULT 6			
RICMSF			
LOCUS			
DEFINITION		Oryza sativa DNA fragment with a miscellaneous signal and an open	PLN 16-JUL-1993
ACCESSION	L12252		
VERSION	L12252.1	GI:169804	
KEYWORDS			
SOURCE		Oryza sativa (japonica cultivar-group)	
ORGANISM		Oryza sativa (japonica cultivar-group)	
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
AUTHORS		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
TITLE		Ehrhartioideae; Oryzeae; Oryza.	
REFERENCE		1 (bases 1 to 825)	
AUTHORS		Krishnan,H.B. and Puebpke,S.G.	
TITLE		Nucleotide sequence of an abundant rice seed globulin: homology	
		with the high molecular weight glutelins of wheat, rye and	
		criticale	
JOURNAL		Biochem. Biophys. Res. Commun. 193 (1), 460-466 (1993)	
MEDLINE		93277591	
PUBMED		8503935	
COMMENT		Original source text: Oryza sativa (cultivar Lamont) 20 days	
		post-anthesis endosperm CDNA to mRNA.	
FEATURES		Location/Qualifiers	
source		1..825	
		/organism="Oryza sativa (japonica cultivar-group)"	
		/mol_type="mRNA"	
		/cultivar="Lamont"	
		/db_xref="taxon:39947"	
		/tissue_type="endosperm"	
		/dev_stages="20 days post-anthesis"	
		1..54	
		misc_signal	
		25..549	
		/codon_start=1	
		/label=ORF	
		/protein_id="AAA72362.1"	
		/db_xref="GI:169805"	
		/translation="MAAMVAISGAHVSESMRFRDQCQREVODSPDLDAQVLDRLD	
		TGRERFMFRPCALGLRMQCCLODVSRRCRAAIRMVRSYEBSMPLEQGS	
		SSSEYVGEGSSSEQYGESSEEGYEQQQPGWTRVLRARQAALPSMR	
		VEPQCSIFAAGQI"	
BASE COUNT	165 a 244 c 270 g 146 t		
ORIGIN			


```

KNDISGTIPTSFLATVAVFSVARNHVHQVPPWLGNTLTALDNLNADNIMSGHVPP
ALSKLINRSLTVAIINNLOGLIPVLFNMSLECLNFGSNQSGSLPQDQISMLPNLK
KGFVFNREGQIPASLSNLSLEHLSHGNFRGRI PSNIGOSGLRTVFEVGNELQ
ATESRDWDLTSLANCSLLLVNQLNLSGILPNSIGNLSQKLEGLRGVGNQIAGLI
PTGIRYLKLAIEFADNRTGTIPSDIGKLSNLSKLSLQFQRIYIEGIEPSSIGNLSQ
NLIALSTNLEGISIPATFNGNLTLISLDIASNLSGKIPEEVNRISSIALFLNLSNL
LDGPI8PHIGQLANLAIIDFSSNKLSPINPALSGLSCIALQFHLQNLIGQGTIPKELM
AURGLEBLDLSNLSGPVPEFLESFOLLKLNLSFNHLSGPPVDKGIETNARVISLIT
SNMGLCGPVFPFPTCPSPDKLASHKLLQILVTVAGARILLGVCIARCVNKS
RGADHQQQATPMFQFISYTELHSATDSFSENLVGRGSPGVSIGTSGSNNLITA
AKVLDVQKQATRSISECNALMIRHRLKLVITVCDLSHNGQFALVLEFIPN
GSDMLHLPSTEDFETPNLMQRLNLTALDAEALYLDHIDPPIVHCDVKSNNILL
DMVAHLGTFGLAKIIRAKSKQSLQSDSCVSGIKGTIGYVAPEYGTGTEISVEGDVY
SYGVLLLEMLTGRPTDPFSDTTLNPKVEMACPNLLETMDNIRCNQEPQAVLEL
FAAPVSRGLGLACGSRQRIKMGDDVVKELGAIKQIIMASQNVASWSTKLY"
31751..31774
/repeat_region
/rpt_type=tandem
/rpt_unit=tttc
38555..38618
/transposon="MITE"
<38912..>39586
/product="globulin"
38912..39586
/codon_start=1
/product="globulin"
/db_xref="GI:304211050.1"
/db_xref="GI:304211166"
/translations="MKKFPFAVFLTLTVTISAQGVLSQSLADACRGEVQAKPLLA
CQRIHQHTGAVGVPPQAGARDCCQQLSVSRGCRCSALRGMDVDEYQSMPP
LEGRRRSGRGEQSGESTAEQQEQVGGYSGTGESQOQGGYGHVGTGVRG
GOOQGMCLRRPQRQOQGGFSGEGAQKPKGVRLTKVRLPTACRIPEQBCSVFS
TLPVLG"
complement(41245..42179)
/transposon="Solo-LTR retrotransposon Bare-1_184G9-2a"
complement(42180..42386)
/transposon="LTR retrotransposon Lollaog_184G9-1p"
complement(42387..43152)
/transposon="LTR retrotransposon Bare-1_184G9-3p"
complement(43153..53697)
/transposon="LTR retrotransposon Lollaog_184G9-1p"
complement(53698..53841)
/transposon="Solo-LTR retrotransposon Bare-1_184G9-2a"
53842..58439
/transposon="LTR retrotransposon Leojyg_184G9-1"
complement(58440..59149)
/transposon="Solo-LTR retrotransposon Bare-1_184G9-2a"
<62032..>64305
/product="D-Hordein"
62032..64305
/codon_start=1
/product="D-Hordein"
/db_xref="GI:304211051.1"
/db_xref="GI:304211167"
/translations="MAKRLVLFVAVIVLVALVTLTAREINGNNIFLDSRSRLQOCERE
LOBESLEACRVVDQOLPWSGLQMCQCRDVSPECRPVALSQVVRQYEQOT
EVPKSGSVFVGTAAPLOQGWMTGSKVYVDPDTSQSGQGGYHQVTSQOP
GQOQSGYSGTSPQPGQGPQCPQSPYSATFPQPGQGGYGPATSLQ
PQGGQGPYTSATSPQPGQGGYQBPPIATSPHQPGQGGYGPATSPHQPGQGGY
SQGGQGGYTSATSPQPGQGGYQBPPIATSPHQPGQGGYGPATSPHQPGQGGY
SPQOQGGYQBPPIATSPQPGQGGYQBPPIATSPHQPGQGGYGPATSPHQPGQGGY
HGQLVQGGYQBPPIATSPQPGQGGYQBPPIATSPHQPGQGGYGPATSPHQPGQGGY
SQGVQACQHSSTSPQOQACQASPKQGLSLYPSGAYTOOKPGGVNPGTSP
LHQGGYGGGLTTPQGGYQBPPIATSPHQPGQGGYGPATSPHQPGQGGYGPATSPHQ
QTTVSPHQGGYQBPPIATSPHQPGQGGYQBPPIATSPHQPGQGGYGPATSPHQ
QOQGGYQBPPIATSPHQPGQGGYQBPPIATSPHQPGQGGYGPATSPHQPGQGGY
LPMCRLEGGLASQ"
join(<64782..64942,65077..65414,65452..65555,66035..66224,
66334..66725,66880..>67194)
/product="putative protein kinase"
join(64782..64942,65077..65414,65452..65555,66035..66224,
66334..66725,66880..67194)
/codon_start=1

```

```

/product="putative protein kinase"
/protein_id="AAP31052.1"
/db_xref="GI:304211168"
/translations="MIGTEVLASHSVDPFAIRKVAARGRVVFAAIRWGAARGRVVFA
AVRWGAARGRVVFAIRKVAARGRVVFAIRKVAARGRVVFAIRKVAARGRVVFA
GRGFCMGASHEABEPLTSETKKRKNVAASPEPKRMPKNIWKKKAKNQLATLV
KEISLTKNSAMHVNILNMINICTPTQYEEHSYLRNLAVIAVKQLDKDGLQGNRE
FLVEVLMISLHHPNLVLLGYCTECQDKILVYEMPLGSLQDLHLLDLPKSOPLSWH
TRMKTAAADAARGLEVLEHVPVVRDLKASNILLDGNFSAKLADFLAKLPVGDK
THVTRVWGTYCYCAPEYAMSKLTKMSDIYCFGVVLELITGRRAIDTTKPTREOIL
VHWAAPLFKDKKTKMADPLDSDSKTPKGLYQALASISMCLOBERAISRLISDVVTA
LTFLLADPNYDPPDIPIPIPVNPDGKISLREABISLISGFFBEKQVEDS"
70339..77371
/transposon="non-LTR retrotransposon Leva_184G9-1"
70469..70482
/rpt_type=tandem
/rpt_unit=ta
complement(77378..>77855)
/transposon="LTR element Wobi_184G9-1p"
77856..86572
/transposon="LTR retrotransposon Bare-1_184G9-4"
complement(<86571..87579)
/transposon="LTR element Wobi_184G9-1p"
87266..87285
/rpt_type=tandem
/rpt_unit=ag
complement(87580..89305)
/transposon="LTR retrotransposon solo-LTR Bare-1_184G9-5a"
complement(<89306..90807)
/transposon="LTR element Wobi_184G9-1p"
complement(<91388..>91975)
/transposon="LTR retrotransposon Wham_184G9-1p"
complement(<93655..>95251)
/transposon="LTR retrotransposon Barbara_184G9-1p"
95284..>98782
/transposon="LTR retrotransposon Uaier_184G9-2p"
complement(<98783..100481)
/transposon="LTR retrotransposon Bare-1_184G9-6p"
100482..101938
/transposon="LTR retrotransposon solo-LTR Wham_184G9-2a"
101839..109132
/transposon="LTR element Egug_184G9-1"
109133..109156
/transposon="LTR retrotransposon solo-LTR Wham_184G9-2a"
complement(109157..>109388)
/transposon="LTR retrotransposon Bare-1_184G9-6p"
complement(109389..119184)
/transposon="LTR retrotransposon Wham_184G9-3"
complement(<119185..120562)
/transposon="LTR retrotransposon Bare-1_184G9-6p"
BASE COUNT 32679 a 27422 c 28340 g 32121 t
ORIGIN
Alignment Scores:
Pred. No.: 3,63e-10 Length: 120562
Score: 381.00 Matches: 98
Percent Similarity: 51.72% Conservative: 22
Best Local Similarity: 42.24% Mismatches: 50
Query Match: 33.19% Indels: 62
DB: 8 Gaps: 10
US-10-053-410-4 (1-206) x AY268139 (1-120562)
QY 11 AlaLeuCysePhaAlaLeuValAlaValAlaValCysGlnGlyClnuValGluArgGln 30
Db 38933 GCGGTGTTCTCAGCACCCCTGTCACCATCTCCGCCGCCCAAGCGGTGTCAG---CAG 38989
QY 31 ArgLeuArgAspLeuGlnCystpGlnGluGluGluSerProteinAspAlaCysArg 50
Db 38990 AGCCTCGCGGACGCCGAGTGCCTCGCGGCGGAGGTCAGGCGAAGCCGCTCTCGGTCGCG 39049
QY 51 GlnValLeuAspArgGlnLeuThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


Db 439 CAACAGCCAGGACAAGGGCAACAACCTGGCCAATATGGCAGAAGAACAGGACAAAGGCCAACA 498

Qy 134 TyrTyrrProCysser-----ArgProGluGlyTyrrGlyTyrrGlyGlnGlyGly 151
:::|||||:::|||||:::|||||:::|||||

Db 499 TGGTACTACCNACTTCTTCGCAGCAGCCAGGACAAGGGCAACAGATAGGAAGGG--- 555

Qy 152 GlnArgGlnMetTyrrProProCyArGProGlyThrThrGlyGlyGlyProArgIleGly 171
:::|||||:::|||||:::|||||:::|||||

Db 556 ---CAAACAAGGGTACTACCCAACTCTCTGCAGCAGCGAGGACAAGGGCAACAATAAGGA 612

Qy 172 Arg 172

Db 613 CAA 615
:::

RESULT 10
TAGLUIDG 3095 bp DNA linear PLN 12-SEP-1993
LOCUS Wheat gene for HMW-glutenin subunit from chromosome 1D.
DEFINITION X03041
ACCESSION X03041.1 GI:21778
VERSION
KEYWORDS gluten; glutenin; seed storage protein; storage protein.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 3095)
AUTHORS Thompson,R.D., Bartels,D. and Harberd,N.P.
TITLE Nucleotide sequence of a gene from chromosome 1D of wheat encoding
a HMW-glutenin subunit
JOURNAL NuclAcid Acids Res. 13 (19), 6833-6846 (1985)
MEDLINE 86041882
PubMed 3840588
COMMENT The central repetitive region between pos. 801 and 2270 is composed
of two multiple repeated amino acid motifs: PQQGQ and GYPTSLQQ.
FEATURES
source 1..3095
/organism="Triticum aestivum"
/mol_type="genomic DNA"
/db_xref="taxon:4565"
misc_feature 202..218
/note="sequence homologous to 5' end of a gliadin gene"
promoter 334..340
/note="put. TATA-box"
CDS 426..2408
/note="LDy-HMW-glutenin (aa 1-660)"
/codon_start=1
/protein_id="CAA26847.1"
/db_xref="GI:21779"
/translation="MAKRLVFAAVIALVALTAEGASRQLQCERLEOBSSEILAECR
OVVDQLAGRLPSTGLMRCCLRDVSAKRSVAIVSQARQEVTPVPKGGSFYFP
GETTLPQQLQGIFWGTSSTQGYTPSVTPSGSYPGASPPQEQGQPQGWKE
FGQQWMYYQTIGQGGPEQGPQGGYPTSLQPGGQQTIGQQGYTPSPHT
GRQRQPVQGGIQGGPEQGPQGGYPTSLQPGGQQTIGQQGYTPSPHT
YPTSLQPGGQGHYLASQPAQGGHYPSAQPGQGGHYPASQPGQGGHYPS
GHYPASQPGGQGGQIPASQPGQGGHYPSAQPGQGGHYPSAQPGQGGHY
IGPQKQPGQGGQGTQGGPEQGPQGGHYPSAQPGQGGHYPSAQPGQGGHY
TSIQPGQGGHYPSAQPGQGGHYPSAQPGQGGHYPSAQPGQGGHYPSAQ
QVQPGQGGQGGQGGYPTSPQPGQGGQGGHCPMSPTQAQGLGQGGQGG
QVQPGQGGQGGYPTSLQPGGQGGQGGQGGHCPMSPTQAQGLGQGGQGG
AEQQAAPMVAKAQQPATQLPTVCMEGGDALSAQ"
misc_feature 2463..2468
/note="pot. polyA signal"
misc_feature 2513..2518
/note="pot. polyA signal"
misc_feature 2516..2521
/note="pot. polyA signal"
BASE COUNT 1009 a 820 c 735 g 531 t
ORIGIN

Alignment Scores:
Pred. No.: 9.14e-07 Length: 3095

```
REFERENCE 2 (bases 1 to 2163)
AUTHORS Jiang,Y., Sun,M., Zheng,J., Xiao,Y. and Yan,Y.
TITLE Direct Submission
JOURNAL Submitted (03-MAR-2003) Key Lab of Genetics and Biotechnology, China
Biology Department, Capital Normal University, Beijing 100037, China
FEATURES
source
1..2163
/organism="Triticum turgidum subsp. durum"
/mol_type="genomic DNA"
/cultivar="Semitco"
/sub_species="durum"
/db_xref="taxon:4567"
<1..>2163
/gene="Glu-1B"
/allele="Glu-1B"
<1..>2163
/gene="Glu-1B"
/product="HMW glutenin subunit 1By8"
1..2163
/gene="Glu-1B"
/codon_start=1
/product="HMW glutenin subunit 1By8"
/protein_id="AA064642.1"
/db_xref="GI:29150726"
/translation="MAKRLVLFATVITVLTAAGEASRLQCELELESLEACR
QVVDQLAGRLPMSTGLMRCQQLQDVSARKLVAVSVQVRYEQTVVPRKGGFYP
GETPLQQLQVFWGTSTQVGYVPSVSSPQGPYPQASPOQPGQGGQPGKWOE
LGQGGQGYPTSLHOSGQGGQYVPSLQPGQGGQIGQGGQGYPTSLQPGQGGQI
GQGGQGYPTSPHQPQPGQGGQIGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
QLGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
PQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
QPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
OLEGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
GYVSSSLQPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
SPQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
IGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
VNTQGTASPKVAKQPPATQPLIMCRMEGGDALSQ"
BASE COUNT 737 a 612 c 563 g 251 t
ORIGIN

Alignment Scores:
Pred. No.: 7.18e-07 Length: 2163
Score: 289.00 Matches: 85
Percent Similarity: 42.25% Conservative: 24
Best Local Similarity: 32.95% Mismatches: 47
Query Match: 25.17% Indels: 102
DB: Gaps: 13

US-10-053-410-4 (1-206) x AY245797 (1-2163)
QY 17 LeuValAlaValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGln 36
Db 43 CTCGTGGCTCTACTGCTGTAAGTGAGGCGCTCTAGGCAA-----CTACAG 90
QY 37 CysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGln 56
Db 91 TGTGACGGGAGCTCCAGGAGACTCGTTGAGCATGCGCAGAGCTGTGGACCAACAG 150
QY 57 LeuThrGlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArg 76
Db 151 TTG-----GCCGGTGGCTGCCATGGAGCAGCGGGCTCCAG 186
QY 77 MetArgCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArg 96
Db 187 ATGCGATGTCGACGAGCTCCAGAGCTCCAGATGTTAGTGCGCTAAGTGGCGCTCGTCGCGTCAGC 246
QY 97 SerMetValArgGlyTrpGluGlu----AlaMetProLeuGluGlyGlyTrpTrpPro 115
Db 247 CAAGTCGTAGACAAATATGACCAACCGGTGTCGCCGCCCAAGGGCGGATCTCTTACCT 306
QY 116 -----TrpGlyArg----- 118
Db 307 GCGGAGACACACCCTGCAGCAACTCCACAAAGTAATATTTTGGGGACATCTTCACAA 366
```

```
QY 119 -----GlnGlnGlnPro----- 122
Db 367 ACAGTACAGGGTATTACCAAGCGTAAGTTCTCTCAGCAGGGGCCATATTATCCAGGC 426
QY 123 -----ProProGlnGlyGly-----Gly 128
Db 427 CAAGCTTCTCCACAACAGCCAGGCAACAGCCAGGCAACAGGCAAAATGGCAAGAACTGGGA 486
QY 129 GlyGlyGlnGlyGlyTyr----- 134
Db 487 CAAGGGCAACAAGGGGTACTTACCCAACTTCTGTGCATCAGTCAGGCAACAGGCAACAAAGGG 546
QY 135 TyrTyrProCysSer-----ArgProGlyGluGlyTyrGlyTyrGlyGlnGly----- 150
Db 547 TACTACCCATCTTCTGTGAGCAACAGGCAACAGGCAACAGGCAACAGGCAACAA 606
QY 151 -----GlyGln 152
Db 607 GGATACTACCCAACTTCTGTGAGCAACAGGCAACAGGCAACAGGCAACAGGCAAC 666
QY 153 ArgGlnMetTyrPro-----ProCysArgProGly-----ThrThrGlyGlyGly 167
Db 667 CAAGGGTACTACCCAACTTCTGTGAGCAACAGGCAACAGGCAACAGGCAACAA 726
QY 168 ProArgIleGlyArg---ValArgLeuThrLysAlaArgGluTyrAlaAlaGly 184
Db 727 CAGCAATAGGCAACAGGCAACAGGCAACAGGCAACAGGCAACAGGCAACAGG 780
RESULT 12
TAE314785 2330 bp DNA linear PLN 11-JUN-2003
LOCUS Triticum aestivum glu-1R gene for high molecular weight glutenin
subunit Y, clone pHTFasR11.
ACCESSION AJ314785
VERSION AJ314785.1 GI:14329762
KEYWORDS glu-1R gene; high molecular weight glutenin subunit Y.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Triticum.
1
De Bustos A., Rubio P. and Jouve N.
Characterization of two gene subunits on the 1R chromosome of rye
as orthologs of each of the Glu-1 genes of hexaploid wheat
Theor. Appl. Genet. 103, 733-742 (2001)
2
de Bustos A. and Jouve N.
Characterisation and analysis of new HMW-glutenin alleles encoded
by the Glu-R1 locus of Secale cereale
Theor. Appl. Genet. 107, 74-83 (2003)
3 (bases 1 to 2330)
/clone="pHTFasR11"
/chromosome="1RL"
/note="1R addition line obtained from T.aestivum cv.
Chinese Spring and S. cereale cv. Imperial as donor of 1R
chromosome"
140..2302
/gene="glu-1R"
140..2302
/gene="glu-1R"
/codon_start=1
FEATURES
source
1..2330
/organism="Triticum aestivum"
/mol_type="genomic DNA"
/db_xref="taxon:4565"
/chromosome="1RL"
/note="pHTFasR11"
/note="1R addition line obtained from T.aestivum cv.
Chinese Spring and S. cereale cv. Imperial as donor of 1R
chromosome"
140..2302
/gene="glu-1R"
140..2302
/gene="glu-1R"
/codon_start=1
```

```
/evidence=experimental
/product="high molecular weight glutenin subunit y"
/protein_id="CAC40687.1"
/db_xref="GI:14329763"
/db_xref="GOA:Q941J6"
/db_xref="SPTREMBL:Q941J6"
/translations="MAKRLVLFATVITLVALTAAGEASRLQCERELQESSLEACR
QVVDQLAGRLPMSGTGLMRCCQLRDVSAKRLVAVSVVROVEYTVVPPKGSFYP
GETTLQQLQVLFMTSGTVGYVSVSSPOGSPYVPGQASPOGPGQPGKQWE
LGQGGQYPTSLHSGGQGGYPSLSQPGGQQLGGQGGYPTSLQPGGQQL
GGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSLQPGGQQL
OLGQGGQPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
PGQGGQPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
PGQGGQPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
OLPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
OLPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
GIYSSLPQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
SPOGPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
IGQVPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
VNTQQTASPVKVQQPATQLPIMCRMEGGDALSASQ"
140. 202
/sig_peptide
140. 202
/mat_peptide
140. 202
/evidence=experimental
/product="HMW glutenin subunit y"
/evidence=experimental
BASE COUNT 794 a 668 c 583 g 285 t.
ORIGIN

Alignment Scores:
Pred. No.: 7.64e-07 Length: 2330
Score: 289.00 Matches: 85
Percent Similarity: 42.25% Conservative: 24
Best Local Similarity: 32.95% Mismatches: 47
Query Match: 25.17% Indels: 102
DB: 8 Gaps: 13

US-10-053-410-4 (1-206) x TAE314785 (1-2330)
QY 17 LeuValAlaValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGln 36
DB 182 CTGCTGGCTCTACTGCTGCTGAGGTGAGGCTCTAGGCAA-----CTACAG 229
QY 37 CysTrpGlnValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGln 56
DB 230 TGTAGCGCGAGCTCCAGGAGAGCTCGCTTGGAGCATGCCGAGGTGCTGGACCAACAG 289
QY 57 LeuThrGlyGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGlyLeuArg 76
DB 290 TTG-----GCCGTGGCTGCTGCTGAGGAGGAGGCTCCAG 325
QY 77 MetArgCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaLeuArg 96
DB 326 ATGCGATGCTGCCAGCAGCTCCAGATGTTAGCGCTAAGTGCCTCTCGTCCGCGTCAGC 385
QY 97 SerMetValArgGlyTrpGluGlu---AlaMetProLeuGluLysGlyTrpTrpPro 115
DB 386 CAAGTCGTAGACAAATATGAGCAACCGTGTGCGCCGCCAAGCGCGGATCTCTTACCCCT 445
QY 116 -----TrpGlyArg----- 118
DB 446 GCGGAGACACACCCTGCAGCACTCCACAGTAATATTTTGGGGAGCATCTTCACAA 505
QY 119 -----GlnGlnPro----- 122
DB 506 ACAGTACAGGGTATTACCACAGGTAAGTTCCTCCTCAGCAGCGGGCCATATTATCCAGGC 565
QY 123 -----ProGlnGlyGly-----Gly 128
DB 566 CAAGCTTCTCCACAGCAGCAGGCAAGGCAACAGCAGGCAATGGCAAGAACTGGGA 625
QY 129 GlyGlyGlnGlyGlyTrp----- 134
```

```
Db 626 CAAGGCAACAAGGGTACTACCCAACTTCTGTCATCAGTCAGGACAAGGACAACAAGG 685
QY 135 TyrTyrProCysSer-----ArgProGlyGluGlyTyrGlyTyrGlyGlnGly----- 150
Db 686 TACTACCCATCTTCTCTGTCAGCAACAGGACAAGGGCAACAGATAGGACAAGGGCAACA 745
QY 151 -----GlyGln 152
Db 746 GGATACTACCACTTCTCTGTCAGGAGGACAAGGCAACAGATAGGACAAGGACA 805
QY 153 ArgGlnMetTyrPro-----ProCysArgProGly-----ThrThrGlyGlyGly 167
Db 806 CAAGGTACTACCACTTCTCTGTCAGCAACAGGACAAGGGCAACAGATAGGACAAGG 865
QY 168 ProArgIleGlyArg---ValArgLeuThrLysAlaArgGluTyrAlaAlaGly 184
Db 866 CAGCAATAGGACAAGGGCAACAGTACGACAAGGGCGGCAATAGGACAAGG 919

RESULT 13
AY263343
LOCUS
DEFINITION
Thinopyrum elongatum clone pUC2b high molecular weight glutenin
y-subunit gene, complete cds.
ACCESSION
AY263343
VERSION
AY263343.1 GI:30983896
KEYWORDS
Thinopyrum elongatum
ORGANISM
Thinopyrum elongatum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Thinopyrum.
REFERENCE
1 (bases 1 to 1980)
AUTHORS
Xia,G.M., Feng,D.S. and Chen,F.G.
TITLE
Agropyron elongatum high molecular weight glutenin y-subunit gene
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 1980)
AUTHORS
Xia,G.M., Feng,D.S. and Chen,F.G.
TITLE
Direct Submission
JOURNAL
Submitted (25-MAR-2003) Plant Cell Engineering Laboratory, School
of Life Science, Shandong University, Shan Da Nan Road 27, Jinan,
Shandong 250100, China
FEATURES
Location/Qualifiers
1..1980
/organism="Thinopyrum elongatum"
/mol_type="genomic DNA"
/db_xref="taxon:4588"
/clone="pUC2b"
1..1977
/codon_start=1
/product="high molecular weight glutenin y-subunit"
/protein_id="AAP41067.1"
/db_xref="GI:30983897"
/translation="MAKRLVLFATVITLVALTAAGEASRLQCERELQESSLEACR
QVVDQLAGRLPMSGTGLMRCCQLRDVSAKRLVAVSVVROVEYTVVPPKGSFYP
GETTLQQLQVLFMTSGTVGYVSVSSPOGSPYVPGQASPOGPGQPGKQWE
LGQGGQYPTSLHSGGQGGYPSLSQPGGQQLGGQGGYPTSLQPGGQQL
GGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSLQPGGQQL
OLGQGGQPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
PGQGGQPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
PGQGGQPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
OLPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
OLPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
GIYSSLPQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
SPOGPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
IGQVPGQGGQGGYPTSPHQRQPGQGGYPSLSQPGGQQLGGQGGYPTSP
VNTQQTASPVKVQQPATQLPIMCRMEGGDALSASQ"
BASE COUNT 654 a 577 c 524 g 225 t
ORIGIN

Alignment Scores:
Pred. No.: 8.46e-07 Length: 1980
Score: 287.00 Matches: 85
Percent Similarity: 42.80% Conservative: 22
Best Local Similarity: 34.00% Mismatches: 47
Query Match: 25.00% Indels: 96
```

[illegible]

```

Db 937 ATGGATGCTGCCAGCAGCTCCGAGATGTTAGCGTAAGTGCCTCCCGTCGCGTCAGC 996
QY 97 SerMetValArgGlyTyrGluGlu----AlaMetProProLeuGluGlyGlyTyrTrpPro 115
Db 997 CAAGTCGTAGACAATATAGCAACACCGTGGTCCGCCCAAGGCGGATCCTTCTACCT 1056
QY 116 -----TrpGlyArg-
Db 1057 GCGAGACACACACTGCAGCACTCCAAACAAGTAATATTTGGGGAACATCTTCACAA 1116
QY 119 -----GlnGlnPro-
Db 1117 ACAGTACAAGGGTATTACCAAGCGTAAGTCTCTCAGCAGGGGCCATATTATCCAGGC 1176
QY 123 -----ProProGlnGlyGly-
Db 1177 CAAGCTTCTCCACAACAGCAGCAGGCAAGGCAACAGCCAGGCAAAATGCCAAGAACTGGGA 1236
QY 129 GlyGlyGlnGlyGlyTyr-
Db 1237 CAAGGGCAACAAGGGTACTACCAACTTCTCTGCATCAGTCAGGACAAGGACAACAAGG 1296
QY 135 TyrTyrProCysSer-----ArgProGlyGluGlyTyrGlyTyrGlyGlnGly- 150
Db 1297 TACTACCATCTCTCTGCAGCAACAGGCAAGGCAACAGATAGGACAAGGCAACAA 1356
QY 151 -----GlyGln 152
Db 1357 GGATACTACCAACTTCTCTGCAGCAGCAGGCAAGGCAACAGATAGGACAAGGACAA 1416
QY 153 ArgGlnMetTyrPro-----ProCysArgProGly-
Db 1417 CAAGGGTACTACCAACTTCTCGCAACACCCAGGCAAAAGGCAACCAAGGCAAGG 1476
QY 168 ProArgIleGlyArg---ValArgLeuThrLysAlaArgGluTyrAlaAlaGly 184
Db 1477 CAGCAATAGGACAAAGGCAACAACTAGGACAAGGCGGCAATAGGACAAGG 1530

RESULT 15
AY263345
LOCUS
DEFINITION
Thinopyrum elongatum clone pUCG7a high molecular weight glutenin
Y-subunit gene, complete cds.
ACCESSION
AY263345
VERSION
AY263345.1 GI:30983900
KEYWORDS
Thinopyrum elongatum
SOURCE
Thinopyrum elongatum
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Thinopyrum.
REFERENCE
1 (bases 1 to 906)
Xia,G.M., Feng,D.S. and Chen,F.G.
Agropyron elongatum high molecular weight glutenin y-subunit gene
Unpublished
2 (bases 1 to 906)
Xia,G.M., Feng,D.S. and Chen,F.G.
Direct Submission
Submitted (26-MAR-2003) Plant Cell Engineering Laboratory, School
of Life Science, Shandong University, Shan Da Nan Road 27, Jinan,
Shandong 250100, China
FEATURES
source
1. 906
/organism="Thinopyrum elongatum"
/mol_type="genomic DNA"
/db_xref="taxon:4588"
/clones="pUCG7a"
1. 903
/codon_start=1
/product="high molecular weight glutenin y-subunit"
/protein_id="AAP41069.1"
/db_xref="GI:30983901"

CDS

```

```

/translation="MAKRLVLPATVVLALMALTAAGEASRLQCRERELESLEACR
QVVDQLAQLPWSTGLMRCCQQLRDVSAKRLIAVSOVARQYEQOTAVPKGGPFY
PGETAPPOLOQRIEFGWSSETVQGYPSVTSYQPGKASYPGKASPOPGQGGQGMQ
EFGCQGGYPTSLQPGQGGKARSPPHSCWQCAGWSGATHCRFGQGGQGYPTSLQ
PQGGQGGKGGQGGHQPQGGHQPQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG
AQQPATQLSAMCRMEGGDALSQ"
BASE COUNT 243 a 265 c 275 g 123 t
ORIGIN
Alignment Scores: 4,94e-07 Length: 906
Pred. No.: 286.00 Matches: 87
Score: 38.61% Conservative: 30
Percent Similarity: 38.61%
Best Local Similarity: 28.71% Mismatches: 50
Query Match: 24.91% Indels: 136
DB: 8 Gaps: 12
US-10-053-410-4 (1-206) x AY263345 (1-906)
QY 16 AlaLeuValAlaValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeu 35
Db 40 GCCTCATGGCTCTACCCGCGCTGAAGGTGAAGCCTCTAGGCNA-----CTA 87
QY 36 GlnCysTrpGlnGlnValGlnGlnSerProLeuAspAlaCysArgGlnValLeuAspArg 55
Db 88 CAGTGTAGCGCGAGCTCCAGGAGAGCTCGCTTGAGCATGCCGCGAGGTCGTGGACCA 147
QY 56 GlnLeuThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 75
Db 148 CAGTTG-----GCTGGCCAGCTGCTCATGAGCAGCGGGCTC 183
QY 76 ArgMetArgCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIle 95
Db 184 CAGATCGGTGTCTCCAGCAGCTCCAGATGTTAGCCCAAGTGGCGCTCATCGCGTC 243
QY 96 ArgSerMetValArgGlyTyrGluGlu-----AlaMetProProLeuGluLysGlyTrp 113
Db 244 AGCCAAGTCGCAAGACAATACGAGCAACAACCGCGGTGCGGCCCAAGCGGCCCTTC 303
QY 114 TrpPro-----TrpGlyArgGln 119
Db 304 TACCCCGGGAGAGACCGCGCCACCGCAGCAACTCCAAACAAGAATATTTTGGGGAAGATCT 363
QY 119 ----- 119
Db 364 TCAGAAACAGTACAGGGTATTACCAAGGTAATCTTCTCTCAGCGGGGTCTATATTAT 423
QY 120 -----GlnGlnPro- 122
Db 424 CCAGGCAAAAGCTTCTCCACACACAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 483
QY 123 -----ProProGlnGlyGly 127
Db 484 CCAGGCAAAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 543
QY 127 ----- 127
Db 544 AAAGCGCAGCAGCCCTCGACACAGCTGTTGGCAATGTCCGGATGGAGTGGGGGCGACG 603
QY 128 -----GlyGlyGlnGlyGlyTyrTyrTyrProCysSer-----ArgPro 141
Db 604 CATTGTGGCAGGACAAGGGCAACAGGG-----TACTACCAACTTCTTTGCGAGCAGCA 660
QY 142 GlyGluGlyTyrGlyTyrGlyGlnGlyGln-----ArgGlnMetTyrProProCysArg 160
Db 661 GGACAAGGGCAACAGTCAGGAAAGGGCAACACAGGCAAGGCAAGGCAAGGCAAGGCAAG 720
QY 161 ProGlyThrThrGlyGlyProArgIleGlyArg----- 172
Db 721 GGACACCGCAGGACAAGGGCAGCAATCAGGACAAGGCAAGGCAAGGCAAGGCAAGGCAAG 780
QY 173 -----ValArgLeuThrLysAlaArgGluTyr 181

```


Db 781 TACCATGTTAGCGCGAGCAGCGCGCGCGCTAAAGGTGGCAAAGCGCGCAGAGCCC 840
Qy 182 AlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePheSer 201
Db 841 CCGACACAGCTGTGCGCAATGTGCGGATG-----GAG 873
Qy 202 GlyGlyAsp 204
Db 874 GGGGGCGAC 882

Search completed: November 29, 2003, 18:16:13
Job time : 3585 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 29, 2003, 17:10:31 ; Search time 2046 Seconds
(without alignments)
2447.078 Million cell updates/sec

Title: US-10-053-410-4
Perfect score: 1148
Sequence: 1 MAKIAAAAAALCFALVAV.....NMCRILSEPOECIFSGGDQY 206

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-Q/cgn2_1/USPTO.spool/US10053410/runat_28112003_140757_27896/app_query.fasta_1.391
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10053410 @CGN 1 1 2810 @runat_28112003_140757_27896 -NCFU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pin.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1148	100.0	900	11	AY105070	Zea mays
c	1148	100.0	935	29	CC349094	OGURN267H
3	1122	97.7	1038	29	CC349105	CC349105 OGURN267V
c	857.5	74.7	674	14	CA401604	CA401604 EL01N0423
5	815.5	71.0	602	9	AI711821	AI711821 605064E02
6	772	67.2	586	10	BE917827	BE917827 OV1_7_F02
c	753.5	65.6	605	10	BE917909	BE917909 OV1_7_F02
8	746	65.0	539	9	AI745997	AI745997 605078E03
c	704	61.3	473	14	CD443922	CD443922 EL01N0432
10	625	54.4	560	9	AI712201	AI712201 605065F05
c	615	53.6	570	14	CA402183	CA402183 EL01N0432
12	547.5	47.7	425	10	EG048804	EG048804 OV1_23_C0
13	457.5	39.9	496	10	BG049097	BG049097 OV1_23_C0
c	432.5	37.7	683	9	AU164031	AU164031 AU164031
15	431	37.5	389	9	AI670620	AI670620 605035H03
c	431	37.5	712	9	AU163963	AI63963 AU163963
17	425.5	37.1	595	13	BQ172799	BQ172799 1008G04_X
18	425.5	37.1	704	9	AU094496	AU094496 AU094496
19	422.5	36.8	697	9	AU094596	AU094596 AU094596
20	421.5	36.7	756	13	BQ06521	BQ06521 WHE3580_A
c	418	36.4	707	9	AU164041	AU164041 AU164041
22	415.5	36.2	687	9	AU163861	AU163861 AU163861
23	413.5	36.0	624	9	AU163975	AU163975 AU163975
24	413.5	36.0	687	9	AU163860	AU163860 AU163860
25	413	36.0	706	9	AU163839	AU163839 AU163839
26	411.5	35.8	677	9	AU094642	AU094642 AU094642
27	410.5	35.8	701	9	AU163968	AU163968 AU163968
c	409.5	35.7	710	9	AU164043	AU164043 AU164043
29	409.5	35.7	711	9	AU094571	AU094571 AU094571
30	409.5	35.7	711	9	AU094635	AU094635 AU094635
31	408.5	35.6	705	9	AU165812	AU165812 AU165812
32	407.5	35.5	527	9	AU076079	AU076079 AU076079
33	406.5	35.4	698	9	AU164035	AU164035 AU164035
34	406	35.4	707	9	AU164016	AU164016 AU164016
35	404.5	35.2	656	9	AU094570	AU094570 AU094570
36	404.5	35.2	688	9	AU163853	AU163853 AU163853
37	404.5	35.2	712	9	AU094563	AU094563 AU094563
38	402.5	35.1	650	9	AU091910	AU091910 AU091910
39	402	35.0	652	14	CA402267	CA402267 EL01N0434
40	401.5	35.0	675	9	AU094637	AU094637 AU094637
41	401	34.9	367	12	BM500544	BM500544 PAC000000
42	400.5	34.9	661	9	AU165797	AU165797 AU165797
43	399.5	34.8	643	9	AU164000	AU164000 AU164000
44	399.5	34.8	669	9	AU163999	AU163999 AU163999
45	399	34.8	708	9	AU163878	AU163878 AU163878

ALIGNMENTS

RESULT	1	900 bp	mRNA	linear	HTC	16-OCT-2002
LOCUS	AY105070	PCO101751	mRNA sequence.			
DEFINITION	Zea mays					
ACCESSION	AY105070					
VERSION	AY105070.1	GI:21208148				
KEYWORDS	HTC.					
SOURCE	Zea mays					
ORGANISM	Zea mays					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.					
REFERENCE	1 (bases 1 to 900)					

AUTHORS Hayney, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M., and Tinsley, S.V.

TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes

JOURNAL Unpublished (2002)

REFERENCE 2 (bases 1 to 900)

AUTHORS Coe, E.H.

TITLE Direct Submission

JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES Location/Qualifiers

source

1..900

/organism="Zea mays"

/mol_type="mRNA"

/db_xref="maizeDB:635853"

/db_xref="taxon:4577"

/clone_lib="Maize Mapping Project/DuPont Consensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 191 a 244 c 331 g 134 t

ORIGIN

Alignment Scores:

Pred. No.: 4.92e-73 Length: 900

Score: 1148.00 Matches: 206

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 11 Gaps: 0

US-10-053-410-4 (1-206) x AV105070 (1-900)

QY 1 MetAlaValLeuAlaAlaAlaAlaAlaLeuCysPheAlaAlaValAlaVal 20

Db 68 ATGGCTAAGATCGCCGGCGGGCGGCGCTGTGCTTCGGCGCTTGGTGGCGGTG 127

QY 21 AlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40

Db 128 GCCGTCTGCCAAGCGAGGTGAGCGCGAGGCTCAGGACCTCAGGTGCTGGCAGGAG 187

QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 60

Db 188 GTCCAGGAGACCGCTCGACGCTGCCCGCAGGTCTTCGCGCGAGCTAACCGCGGC 247

QY 61 GlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCys 80

Db 248 GCGCGCGCGCGCGCTGGCCCGCTTCGCTGGCGCACCGCGCTCGGATCGCGTGC 307

QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100

Db 308 CAGCAGCTCCAGGACGTGAGCGCGAGTGGCTGCGCGCCCATCCGAGCATGGTCAGG 367

QY 101 GlyTyrGluGluAlaMetProProLeuGluGlyGlyTyrTrpTrpTrpGlyArgGlnGln 120

Db 368 GGCTACGAGGAGCCATCGCCCGCTGAGAAAGGTGGTGGCATGGGGCGGCGAGCAG 427

QY 121 GlnProProGlnGlyGlyGlyGlyGlnGlyGlyTyrTyrTyrProCysSerArg 140

Db 428 CAGCCCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 487

QY 141 ProGlyGluGlyTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProCysArg 160

Db 488 CCAGGAGAGGGATACGGCTACGGTCAGGTGAGGTGCCAGCGGCAGATGTATCCACCGTGTCT 547

QY 161 ProGlyThrGlyGlyProArgIleGlyArgValArgLeuThrLysAlaArgGlu 180

Db 548 CCGCGCACCCAGCGCGCGGCAAGATCGCGCGCGGTACGAGGCTTACGAGGCGCGGAG 607

QY 181 TyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePhe 200

Db 608 TACGCCCGCGGGGTGCCGATGATGTCCGGCTGTCCGAGCCCGAGAGTGCAGCATCTTC 667

QY 201 SerGlyGlyAspGlnTyr 206

Db 668 TCCGGCGCGGACCATGATAC 685

RESULT 2

CC349094/c

LOCUS

DEFINITION

OGUEN26TH ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0422E04,

genomic survey sequence.

ACCESSION

CC349094

VERSION

CC349094.1 GI:30818501

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 935)

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Numberg, A., Robbins, D. and Lakey, N.

TITLE

Consortium for Maize Genomics

JOURNAL

Unpublished

COMMENT

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

FEATURES

Location/Qualifiers

source

1..935

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBMA0422E04"

/clone_lib="ZM_0.7-1.5 KB"

/note="Vector: pBCKS-; Site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"

BASE COUNT 147 a 339 c 259 g 190 t

ORIGIN

Alignment Scores:

Pred. No.: 5.12e-73 Length: 935

Score: 1148.00 Matches: 206

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 29 Gaps: 0

US-10-053-410-4 (1-206) x CC349094 (1-935)

QY 1 MetAlaValLeuAlaAlaAlaAlaAlaLeuCysPheAlaAlaValAlaVal 20

Db 834 ATGGCTAAGATCGCCGGCGGCGCGGCGGTGTGCTTCCGGCCCTTGGTGGCGGTG 775

QY 21 AlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40

Db 774 GCCGTCTGCCAAGCGGAGGTTCGAGCGGCGAGGCTCAGGACCTGCAGTGTCTGGCAGGAG 715

```

QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 60
Db 714 GTCCAGAGAGCCCGCTCGACGCTGCCCGCGAGCTCTGACCGGAGCTAACCGCGCGC 655

QY 61 GlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCys 80
Db 654 GCGCGCGCGCGCGCGCTTGGCCGTTCCGGTGGGGGACCGCGGCTCCGGATGGGTCTGC 595

QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100
Db 594 CAGCAGCTCCAGGACGTGAGCGCGAGTGCCTGCGCGCCATCCGGAGCATGTGCAGG 535

QY 101 GlyTyrGluGluAlaMetProProLeuGluLysGlyTyrTrpProTyrGlyArgGlnGln 120
Db 534 GGTACAGAGAGCCATGCCCGCTGGAGAAAGGCTGGTGGCCATACCCCTGCAGCCGAG 475

QY 121 GlnProProGlnGlnGlyGlyGlyGlnGlnGlyTyrTyrTyrProCysSerArg 140
Db 474 CAGCCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 415

QY 141 ProGlyGluGlyTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProCysArg 160
Db 414 CCAGGAGAGGATACGGCTACGGTCAAGGTGGCCAGCGGAGATGTATCCACCGTCTGT 355

QY 161 ProGlyThrGlyGlyGlyProArgIleGlyArgValArgLeuThrLysAlaArgGlu 180
Db 354 CCGCGCACCCAGCGCGCGGCGGCAAGGATCGCGCGCGGTGAGGCTTACGAAGGCGCGGAG 295

QY 181 TyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePhe 200
Db 294 TAGCCCGCGGGTGGCCGATGTGTCGGCTGTTCGGAGCCCGGAGGATGCAGCATCTTC 235

QY 201 SerGlyGlyAspGlnTyr 206
Db 234 TCCGGCGCGGACCCAGTAC 217

RESULT 3
CC349105 1038 bp DNA linear GSS 16-MAY-2003
LOCUS OGUN26TV_ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZM58Wa0422E04,
DEFINITION genomic survey sequence.
ACCESSION CC349105
VERSION CC349105.1 GI:30818512
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
AUTHORS A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek
,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Unpublished
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5943
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
FEATURES
    source
        1..1038
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZM58Wa0422E04"
            /clone_lib="ZM_0.7_1.5_KB"
            /notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"

```

```

BASE COUNT 215 a 316 c 336 g 171 t
ORIGIN
Alignment Scores:
Pred. No.: 4,18e-71 Length: 1038
Score: 1122.00 Matches: 206
Percent Similarity: 99.04% Conservative: 0
Best Local Similarity: 99.04% Mismatches: 0
Query Match: 97.74% Indels: 2
DB: 29 Gaps: 0
US-10-053-410-4 (1-206) x CC349105 (1-1038)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaVal 20
Db 393 ATGGCTAAGATCGCCGCGCGCGCGCGCTGTGCTTCGCGGCCCTGGTGGCGGTG 452
QY 21 AlaValCysGlnGlyValGluValGluArgGlnArgLeuArgAspLeuGlnCysTyrGlnGlu 40
Db 453 GCCGTCTGCAAGGCGAGGTCGAGCGGCGCAGAGGCTCAGGAGCTCAGGTGCTGCGCAGGAG 512
QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 60
Db 513 GTCCAGGAGAGCCCGCTCGACGCGTCCGCCAGGTCCTCGACCGGAGCTAACCGCGCGC 572
QY 61 GlyGlyGlyGlyValGlyValGlyProPheArgTyrGlyThrGlyLeuArgMetArgCysCys 80
Db 573 GCGCGCGCGCGCGCGGTGGCCGTTCCGGTGGGGCACCGCGGCTCCGGATGGGTCTGC 632
QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100
Db 633 CAGCAGCTCCAGGACGTGAGCGCGAGTGCCTGCGCGCCATCCGGAGCATGTGTTCAGG 692
QY 101 GlyTyrGluGluAlaMetProProLeuGluLysGlyTyrTrpProTyrGlyArgGlnGln 120
Db 693 GGCTACGAGAGGCGCCATGCCCGCTGGAGAAAGGCTGGTGGCGCATGGGGCGCGCAGCAG 752
QY 121 GlnProProGlnGlnGlyGlyGlyGlnGlnGlyTyrTyrTyrProCysSerArg 140
Db 753 CAGCGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 812
QY 141 ProGlyGluGlyTyrGlyTyrGlyGlnGlnGlyGlnArgGlnMetTyrProCysArg 160
Db 813 CCAGGAGAGGATACGGCTACGGTTCAGGTGCGCGGCGGAGGATGTATCCACCGTCTGT 872
QY 161 ProGlyThrGlyGlyGlyProArgIleGlyArgValArgLeuThrLysAlaArg-Gl 180
Db 873 CCGCGCACCCAGCGCGCGCGGCGGCAAGGATCGCGCGCGGTGAGGCTTACGAAGGCGCGGGA 932
QY 180 uTyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePh 200
Db 933 GTACGCGCGCGGGTGGCCGATGTATGTCGGCTGTTCGGAGCCCGGAGGAGTGCAGCATCTT 992
QY 200 eSerGly-GlyAspGlnTyr 206
Db 993 CTCGCGCGCGCGCAGTAC 1012

RESULT 4
CA401604/c
LOCUS ELO1N0423A08.g EndospERM_4 Zea mays cDNA, mRNA linear EST 07-NOV-2002
DEFINITION
ACCESSION CA401604
VERSION CA401604.1 GI:24766450
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 674)
AUTHORS Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and
Messing,J.
TITLE Sequencing of the maize endospERM ESTs

```



```

Db      189 CCCAGGAGTGACGCTCTCTCCGGCGGACCACTAC 151
RESULT 6
BE917827
LOCUS   BE917827          586 bp      mRNA      linear      EST 29-SEP-2000
DEFINITION OVI_7_F02.g1_A002 Ovary 1 (OVI) sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BE917827
VERSION   BE917827.1 GI:10420322
KEYWORDS Sorghum bicolor (sorghum)
SOURCE   Sorghum bicolor
ORGANISM Sorghum bicolor (sorghum)

REFERENCE
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
TITLE   An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTWix
High quality sequence start: 3
High quality sequence stop: 584
POLYA=No. Location/Qualifiers
source      1..586
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Ovary 1 (OVI)"
/notes="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from Lambda
Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT 115 a 179 c 212 g 80 t
ORIGIN

Alignment Scores:
Pred. No.: 3.07e-46 Length: 586
Score: 772.00 Matches: 144
Percent Similarity: 84.32% Conservative: 12
Best Local Similarity: 77.84% Mismatches: 13
Query Match: 67.25% Indels: 16
DB: 10 Gaps: 4

US-10-053-410-4 (1-206) x BE917827 (1-586)

QY      22 ValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGluVal 41
Db      2 GTCGGCCAGGCGGTGGTCGAGCGGCAGAGGCTCAAGACCTGCGAGTGTCTGGCAAGAGGTC 61
QY      42 GlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 61
Db      62 CAGAGAACCCGCTCGGCGGTGCGGCGAGGTCTGACCGGAGCTCACC---GGCGGC 118
QY      62 GlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGln 81
Db      119 ATCGGCTACGCATCGCCCGCTTCGGTGGGGACCGGGCTCGGATGGGTGCTCCAG 178
QY      82 GlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArgGly 101
Db      179 CAGCTCCAGGACGTGAGCCGCGAGTGCCTGCGCGCCATCCGAGCATGGTGCAGGGGC 238

```

```

QY      102 TYrGluGluAlaMetProProLeuGluGlyGlyTrpTrpProTrpGlyArgGlnGln 121
Db      239 TAGGAGGAGACCATGCCGCGCTGGAGAAAGCTGG-----TGGGGGCGAGCGCCGAG 292
QY      122 ProProProGlnGlnGlyGlyGlyGlnGlyGlyTyTyTyTyTyProCysSerArgPro 141
Db      293 CCG-----GGCTACGACTACCCCTGCAGCCAGCA 322
QY      142 GlyGluGlyTyGlyTyGlyGlnGlnGlyGlnArgGlnMetTyProProCysArgPro 161
Db      323 GGAGAGGGATACGGTACGTCAGAGTGGCCAGCAGCAGATGTACCCACCGTGTCTGCC 382
QY      162 GlyThrThrGlyGlyGlyProArgIleGlyArgValArgLeuThrIlyalaArgGlyTyr 181
Db      383 GGCACC-----GGGCAAAAATCGCCCGGTGAAGCTTACAAAGGCCAGGCAGTAC 433
QY      182 AlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePheSer 201
Db      434 GCCGCGAGGATCCGATGATGTCCGCGTGTCCGAGCCCCAGAGTGCAGCGTCTTCTCC 493
QY      202 GlyGlyAspGlnTyr 206
Db      494 GCGCGCGACCACTAC 508

RESULT 7
BE917909
LOCUS   BE917909          605 bp      mRNA      linear      EST 29-SEP-2000
DEFINITION OVI_7_F02.bl_A002 Ovary 1 (OVI) sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BE917909
VERSION   BE917909.1 GI:10420483
KEYWORDS Sorghum bicolor (sorghum)
SOURCE   Sorghum bicolor
ORGANISM Sorghum bicolor (sorghum)

REFERENCE
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
TITLE   An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 523
POLYA=No. Location/Qualifiers
source      1..605
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Ovary 1 (OVI)"
/notes="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from Lambda
Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT 122 a 184 c 217 g 82 t
ORIGIN

Alignment Scores:
Pred. No.: 6.74e-45 Length: 605
Score: 753.50 Matches: 147

```

```
Percent Similarity: 83.16% Conservative: 11
Best Local Similarity: 77.37% Mismatches: 15
Query Match: 65.64% Indels: 17
DB: 10 Gaps: 5

US-10-053-410-4 (1-206) x BE917909 (1-605)
QY 1 MetAlaLysIleAlaLaA---AlaAlaAlaAlaLeuCysPheAlaAlaLeuValAla 19
Db 82 ATGGCTAAGATCGCTGCGCGCCACCGCGCGCTGCTCGCGCGCTGCTCGCGCGCTGCTGCGC 141
QY 20 ValAlaValCysGlnGlyValGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGln 39
Db 142 GTGGCGCTCGGCCAAGCGTGTCTGCGCGCGCAGAGGCTCAAGAGCCTCGAGTGTGGCAA 201
QY 40 GluValGlnGlnSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly 59
Db 202 GAGGTCAGAGAAACCGCTCGCGCGCTGCGCGCGCTGCTCGCGCGCTGCTCGCGCGCTCAC 258
QY 60 GlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys 79
Db 259 GCGCGATGCGCTACGCGATCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCTGCGCGTGC 318
QY 80 CysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaAlaIleArgSerMetVal 99
Db 319 TGCCAGCAGCTCCAGGAGCGTGAGCGCGAGTGCCTGCGCGCGCTGCGCGCGCTGCGCGTGC 378
QY 100 ArgGlyTrpGluGluAlaMetProProLeuGluGlyGlyTrpTrpProTrpGlyArgGln 119
Db 379 AGGGGCTACGAGGAGACCATGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCG 432
QY 120 GlnGlnProProGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 139
Db 433 CCGCAGCGG-----GGCTACGACTACCCCTGCGAGC 462
QY 140 ArgProGlyGluGlyTrpGlyTrpGlyGlnGlyGlyGlnGlnMetTrpProCys 159
Db 463 CAGCGAGGAGGAGGATACGCTACGCTGAGAGTGCGCGAGGAGGAGGAGGAGGAGGAGGAGG 522
QY 160 ArgProGlyThrThrGlyGlyGlyProArgIleGlyArgValArgLeuThrLysAlaArg 179
Db 523 CGTCCCGGCAC-----GGGCAAAATCGCGCGCTGAAGCTTACAAAGGCGCAGG 573
QY 180 GluTrpAlaAlaGlyLeuProMetMetCys 189
Db 574 CAGTACCGCCAGGATGCGCGATGATGTGC 603

RESULT 8
AI745997/c
LOCUS AI745997 539 bp mRNA linear EST 02-FEB-2000
DEFINITION 605078E03.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
ACCESSION AI745997
VERSION AI745997.1 GI:5124261
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE
AUTHORS Walbot V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL University
COMMENT Contact: Walbot V
Stanford University
Department of Biological Sciences
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605078 row: E column: 03.

FEATURES
source
Location/Qualifiers
1..539
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="PH5(alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: PAD-GAL4-2; Site: 1: EcoRI; Site 2: XhoI; Kernel endosperm cDNA library from Schmidt lab"
BASE COUNT 79 a 204 c 151 g 105 t
ORIGIN
Alignment Scores: 2,07e-44 Length: 539
Pred. No.: 746.00 Matches: 131
Score: 746.00 Conservative: 1
Percent Similarity: 97.78% Mismatches: 1
Best Local Similarity: 97.04% Indels: 2
Query Match: 64.98% Gaps: 1
DB: 9

US-10-053-410-4 (1-206) x AI745997 (1-539)
QY 74 GlyLeuArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAla 93
Db 538 GGGCTCCGATCGGTGTGCTGCCAGCAGCTCCAGGACGTGAGCGCGAGTGCCTCGCTCGGCC 479
QY 94 AlaIleArgSerMetValArgGlyTrpGluGluAlaMetProProLeuGluLysGlyTrp 113
Db 478 GCATCCGAGGATGCTGAGGGGCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 419
QY 114 TrpProTrpGlyArgGlnGlnProProGln-----GlyGlyGlyGlyGlyGln 131
Db 418 TGGCCATGGGGCGCGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 359
QY 132 GlyGlyTrpTrpTrpProCysSerArgProGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 151
Db 358 GGGGGCTACTACTACCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 299
QY 152 GlnArgGlnMetTrpProCysArgProGlyThrThrGlyGlyGlyProArgIleGly 171
Db 298 CAGCGCGCATGTATCCACCTGTGCTCCGCGCACCCCGCGCGCGCGCGCGCGCGCGCGCG 239
QY 172 ArgValArgLeuThrLysAlaArgGluTrpAlaAlaGlyLeuProMetMetCysArgLeu 191
Db 238 CGCGTGAGGCTTACGAAGCGCGGAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 179
QY 192 SerGluProGlnGluCysSerIlePheSerGlyGlyAspGlnTyr 206
Db 178 TCGGAGCCCGCAGGATGCGCTCTTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 134

RESULT 9
CD443922
LOCUS CD443922 473 bp mRNA linear EST 03-JUN-2003
DEFINITION EL01N0432G05.b Endosperm_4 Zea mays cDNA, mRNA sequence.
ACCESSION CD443922
VERSION CD443922.1 GI:31359565
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 473)
AUTHORS Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and
Messing, J.
TITLE Sequencing of the maize endosperm ESTs
JOURNAL Unpublished
COMMENT Contact: Lai, Jinseng
Dr. Joachim Messing's lab
```


Waksman Institute, Rutgers University
130 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@waksman.rutgers.edu
Seq primer: T3.

FEATURES

source

Location/Qualifiers
1. 473
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W22"
/db_xref="taxon:4577"
/tissue_type="Endosperm of 7-23DAP"
/clone_lib="Endosperm 4"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 76 a 138 c 201 g 58 t
ORIGIN

Alignment Scores:

Pred. No.: 1.87e-41 Length: 473
Score: 704.00 Matches: 131
Percent Similarity: 94.96% Conservative: 1
Best Local Similarity: 94.24% Mismatches: 3
Query Match: 61.32% Indels: 4
DB: 14 Gaps: 2

US-10-053-410-4 (1-206) x CD443922 (1-473)

QY 1 MetAlaValyle---AlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAla 19
Db 66 ATGGCTAAGATCGCGCGCGCGCGCGCGCGCTGTGCTTCGGCGCGCTGTGGCC 125
QY 20 ValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGln 39
Db 126 GTGGCCCTCTGCCAAGCGAGGCTGAGCGGAGAGGCTCAGGGCGCGGAGTCTGGCAG 185
QY 40 GluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly 59
Db 186 GAGGGCCAGGAGAGCGCGCTCGACGCTGCGCGAGGCTCTCCACCGGAGCTAAC--- 242
QY 60 GlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys 79
Db 243 -----GGCGCGCGCGCTGCGCGCGCTTCGGTGGGGCACCGGGCTCGGATCGCGTGC 296
QY 80 CysGlnGlnLeuGlnAspValSerArgGluCysArgGluCysAlaAlaAlaLeuSerMetVal 99
Db 297 TGCACAGAGCTCCAGGACGTGAGCGCGAGTGCCTGCGCGCCCATCCGGAGCATGGTC 356
QY 100 ArgGlyTyrgluGluAlaMetProProLeuGluLysGlyTrpTrpProTpglyArgGln 119
Db 357 AGGGCTACAGAGGCGCATCGCGCTCTGAGANAGGCTGTGGCCATGGGGCGGCGAG 416
QY 120 GlnGlnProProGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 138
Db 417 CAGCAGCGCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 473

RESULT 10

AT172201/c

LOCUS

605065F05.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
cDNA, mRNA sequence.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS

TITLE

Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL
COMMENT

University
Unpublished
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605065 row: F column: 05.

FEATURES

source

Location/Qualifiers
1. 560
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site 1: EcoRI;
Site 2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

BASE COUNT 89 a 197 c 159 g 114 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 1.04e-35 Length: 560
Score: 625.00 Matches: 113
Percent Similarity: 89.15% Conservative: 2
Best Local Similarity: 87.60% Mismatches: 12
Query Match: 54.44% Indels: 3
DB: 9 Gaps: 1

US-10-053-410-4 (1-206) x AT172201 (1-560)

QY 80 CysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaAlaLeuSerMetVal 99
Db 560 TGCCACCACTCCAGGACGTACCCCC-GAGTGCCGNTGCCCCCATCCGGAGCATGGTC 502
QY 100 ArgGlyTyrgluGluAlaMetProProLeuGluLysGlyTrpTrpProTpglyArgGln 119
Db 501 AGGGGTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 442
QY 120 GlnGlnProProGln-----GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 137
Db 441 CAGACGCCCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 382
QY 138 CysSerArgProGlyGlyGlyTyrglyTyrglyGlyGlyGlyGlyGlyGlyGlyGlyGly 157
Db 381 TGCAGCGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 322
QY 158 ProCysArgProGlyThrThrGlyGlyGlyProArgIleGlyArgValArgLeuThrLys 177
Db 321 CCTGTGCTCCCGAG 262
QY 178 AlaArgGluTyrgluAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCys 197
Db 261 GCCCGGAGATACCCCG 202
QY 198 SerIlePheSerGlyGlyAspGlnTyr 206
Db 201 AGCGTCTTCGAG 175

RESULT 11

CA402183/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CA402183 570 bp mRNA linear EST 07-NOV-2002
ELOIN0432G05.g Endosperm_4 Zea mays cDNA, mRNA sequence.


```

RESULT 13
BG049097
LOCUS
DEFINITION
OVI_23_C05_g1_A002 Ovary 1 (OVI), Sorghum bicolor cDNA, mRNA
EST 25-JAN-2001
ACCESSION
BG049097
VERSION
BG049097.1 GI:12500510
SOURCE
Sorghum bicolor (sorghum)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 496)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.
An EST database from Sorghum: ovaries of varying immature stages
UNPUBLISHED
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@cuga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTwix
High quality sequence start: 43
High quality sequence stop: 482
POLYA=NO.

FEATURES
source
1..496
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Ovary 1 (OVI)"
/note="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from Lambda
Zap II; Site_1: XhoI; Site_2: EcoRI; the library was made
from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT 125 a 124 c 160 g 87 t
ORIGIN

Alignment Scores:
Pred. No.: 9.63e-24 Length: 496
Score: 457.50 Matches: 84
Percent Similarity: 80.87% Conservative: 9
Best Local Similarity: 73.04% Mismatches: 7
Query Match: 39.85% Indels: 15
DB: 10 Gaps: 3

US-10-053-410-4 (1-206) x BG049097 (1-496)
QY 92 CysAlaAlaIleArgSerMetValArgGlyTyrGluGluAlaMetProLeuGluTys 111
Db 2 TCGCGCGCCATCCGGAGCATGGTCAGGGGCTACGAGGACCATCCGCGCTGGAGAAA 61
QY 112 GlyTyrTrpProTyrArgGlnGlnProProGlnGlyGlyGlyGln 131
Db 62 GGCCTGG-----TGGGGGCGAGCGCGAGCCG----- 88
QY 132 GlyGlyTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlnGlyGly 151
Db 89 ---GGCTACGACTACCCCTCGACCCAGGAGGAGGATACGGCTACGGTGAGAGTGGC 145
QY 152 GlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyProArgIleGly 171
Db 146 CAGCAGCAGATGTACCCACCGTGTCTCCCGGCACC-----GGGCAAAAATCGCC 196
QY 172 ArgValArgLeuThrIysAlaArgGluTyrAlaAlaGlyLeuProMetMetCysArgLeu 191

```

```

RESULT 14
AUI64031
LOCUS
DEFINITION
AUI64031 Rice panicle at ripening stage Oryza sativa (japonica
cultivar-group) cDNA clone E11772, mRNA sequence.
ACCESSION
AUI64031
VERSION
AUI64031.1 GI:11171929
KEYWORDS
EST.
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 (bases 1 to 683)
Sasaki,T. and Yamamoto,K.
Rice cDNA from panicle at ripening stage (2000)
UNPUBLISHED
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP'.

FEATURES
source
1..683
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="E11772"
/dev_stage="ripening stage"
/clone_lib="Rice panicle at ripening stage"
/note="Organ: panicle; Rice cDNA from panicle at ripening
stage"
BASE COUNT 146 a 179 c 242 g 115 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 8.31e-22 Length: 683
Score: 432.50 Matches: 104
Percent Similarity: 63.27% Conservative: 20
Best Local Similarity: 53.06% Mismatches: 44
Query Match: 37.67% Indels: 29
DB: 9 Gaps: 8

US-10-053-410-4 (1-206) x AUI64031 (1-683)
QY 16 AlaLeuValAlaValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeu 35
Db 4 GCCATGGTGGCCATCTCCGGCGCGAGNTGACCGAGTCCGAAGATGAGGTTCAGGGACAGC 63
QY 36 GlnCysTyrGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArg 55
Db 64 CAGTCCACGGGAGGTTCAGGACAGCCCGCTGGACGGTCCGGCAGGTGCTCGACCGG 123
QY 56 GlnLeuThrGly-GlyGlyGlyGlyValGlyValGlyProPheArgTyrGlyThrGlyLe 75
Db 124 CAGCTCACCGCGCGAGAGGTTCAGCGCGATGTTCCCGCGCGCGCGCGC-GCGCTCGCCT 182
QY 75 uArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaI 95
Db 183 CGCGATGTCAGTCTGCCAGCAGCTGCAGACGTGAGCGCGAGTCCGCTGCGCGCCAT 242
QY 95 eArgSerMetValArgGlyTyrGluGluAlaMetPro----ProLeuGluLysGlyTrpTr 114

```

```

243 CCAGCGAGTGTGAGGAGCTACAGGAGAGCATGCCGATGCCCTCGAGCAGAGCTGGTC 302
114 pProTrpGlyArgGlnGlnProProGlnGlnGlyGlyGlyGlyGlyGlyGlyGlyTy 134
303 GTCGTGCTGCTCGAGTACTAC-----GGCGGCGAGGG----- 336
134 rTyTyTyProCysSerArgProGlnGlnGlyTyTyTyTyTyTyTyTyTyTyTyTyTy 154
337 -----TCGTGCTGCGAGCAGGGGTAC---TACGGCGAGGGGTCTCGGAGGA 380
154 nMetTyTy-----ProProCysArgProGlyThrGlyThrGlyGlyProArgIleG 171
381 GGCTACTACGGCAGCAGCAGCAGCAGCGCGGAGTACC----- 420
171 yArgValArgLeuThrLysAlaArgGluTyTyAlaAlaGlyLeuProMetMetCysArg 191
421 -CGCGTGAGGCTGACCCAGGCGAGGAGTACGCGGCGAGCTGCCGTGCGATGTCGGGGT 479
191 userGluProGlnGluCysSerIlePheSerGlyGlyAspGlnTy 206
480 T---GAGCCCCAGCAGTGCAGCATCTTCGCCCGCGGC---CAGTAC 519
```

```

RESULT 15
LOCUS AI670620/c 389 bp mRNA linear EST 02-FEB-2000
DEFINITION 605035H03.x1 605 - EndospERM cDNA library from Schmidt lab Zea mays
cDNA, mRNA sequence.
ACCESSION AI670620
VERSION AI670620.1 GI:4837530
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE 1. (bases 1 to 389)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL Unpublished
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605035 row: H column: 03.
FEATURES
source
1. .389
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endospERM"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5 (alpha)"
/clone_lib="605 - EndospERM cDNA library from Schmidt lab"
/notes="Organ: Kernel; Vector: PAD-GAL4-2; Site 1: EcoRI;
Site 2: XhoI; Kernel endospERM cDNA library from Schmidt
lab"
```

```

BASE COUNT 67 a 136 c 100 g 85 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 6e-22 Length: 389
Score: 431.00 Matches: 76
Percent Similarity: 96.25% Conservative: 1
Best Local Similarity: 95.00% Mismatches: 3
Query Match: 37.54% Indels: 0
DB: 9 Gaps: 0
```

```

US-10-053-410-4 (1-206) x AI670620 (1-389)
QY 127 GlyGlyGlyGlyGlnGlyGlyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 146
DB 387 GGAGGAGGAGGACAGGGGGGCTACTACTACCCCTTCAGCCCGCCAGGAGGATGCCGC 328
QY 147 TyTGlyGlnGlyGlyGlnArgGlnMetTyTyProProCysArgProGlyThrThrGly 166
DB 327 TACGGTCAGGGTGGCCAGCGGCAGATGATATCCACCTGTCTGCCCGGACACCCGGCGGC 268
QY 167 GlyProArgIleGlyArgValArgLeuThrLysAlaArgGluTyTyAlaAlaGlyLeu 186
DB 267 GGSCCAAGGATCGCGCGGTGAGGCTTNCGAAGGCCCGGGAGTACGCCCGGGTTGCCG 208
QY 187 MetMetCysArgLeuSerGluProGlnGluCysSerIlePheSerGlyGlyAspGlnTy 206
DB 207 ATGATGTGCGGCTGTCTCGAGGCCCGCAGGAGTGCAGCGTCTTCTCCGGCGCGCAGGTAC 148
```

Search completed: November 29, 2003, 18:49:18
Job time : 2052 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 29, 2003, 17:10:55 ; Search time 62 Seconds
(without alignments)
1466.531 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1146

Sequence: 1 MAKIAAAAAAALCFALVAV.....MMCRSLSEPFQCSIFSGDQY 206

Scoring table: BLOSUMP62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DB=xlh

-Q/cgn2_1/USPTO_spool/US10053410/runat_28112003.140757.27908/app_query.fasta_1.391
-DB-issued Patents NA -Qfmt=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10053410 @CGN 1.1 56 @runat_28112003.140757.27908 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B COMB.seq:*
5: /cgn2_6/prodata/2/ina/PTUS COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	281.5	24.5	2296	US-08-899-336-2	Sequence 2, Appli
2	130.5	11.4	477	US-08-706-391B-6	Sequence 6, Appli
C 3	124	10.8	554	US-09-252-991A-8810	Sequence 8810, Ap
4	124	10.8	1233	US-09-252-991A-9241	Sequence 9241, Ap
5	124	10.8	1848	US-09-252-991A-9071	Sequence 9071, Ap
C 6	124	10.8	1980	US-09-252-991A-8712	Sequence 8712, Ap
7	123	10.7	2824	US-09-010-928B-3	Sequence 3, Appli
C 8	118.5	10.3	5822	US-08-899-595-4	Sequence 4, Appli
9	118.5	10.3	5822	US-08-899-595-5	Sequence 5, Appli
10	117	10.2	1251	US-09-252-991A-2647	Sequence 2647, Ap
C 11	117	10.2	1323	US-09-252-991A-3022	Sequence 3022, Ap
12	117	10.2	1371	US-09-252-991A-2846	Sequence 2846, Ap

ALIGNMENTS

RESULT 1

US-08-899-336-2
Sequence 2, Application US/08899336
Patent No. 5955649

GENERAL INFORMATION:

APPLICANT: HIROTA, NAHICO

APPLICANT: KIHARA, MAKOTO

APPLICANT: KURODA, HISAO

APPLICANT: ITO, KAZUTOSHI

TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA,

TITLE OF INVENTION: EXPRESSION CASSETTE, EXPRESSION VECTOR AND TRANSGENIC

TITLE OF INVENTION: PLANT

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,336

FILING DATE: 23-JUL-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP HEI 8-193433

FILING DATE: 23-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.

REGISTRATION NUMBER: 24,618

issued 7/29/99

102(6) on "46"

REFERENCE/DOCKET NUMBER: 2589-0061-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2296 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-899-336-2

Alignment Scores:

Pred. No.: 9.58e-14 Length: 2296

Score: 281.50 Matches: 79

Percent Similarity: 46.12% Conservative: 22

Best Local Similarity: 36.07% Mismatches: 51

Query Match: 24.52% Indels: 67

DB: 2 Gaps: 11

US-10-053-410-4 (1-206) x US-08-899-336-2 (1-2296)

QY 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20

DB 37 ATGGCTAAGCGGTGGTCTCTTTGTGGCGGTAAATC-----GTCGCCCTCGTGGCTCTC 90

QY 21 AlaValCysGlnGlyGluValGlu-----ArgGlnArgLeuArgAsp 34

DB 91 ACCACCGCTGAACGTCAGATCAATGGAACAAACATTTCCTTGATAGCGCTCTAGGCAG 150

QY 35 LeuGlnCysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAsp 54

DB 151 CTACAGTGTGAGCGGAGCTCCAGGAGAGCTCGCTGAGCGGTGCGCGGGTCTGGAC 210

QY 55 ArgGlnLeuThrGlyGlyGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGly 74

DB 211 CAACAGCTG-----GTTGGCAGCTGCATGGAGCAGCGGG 246

QY 75 LeuArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAla 94

DB 247 CTCAGATGAGTGTCTGCCAGCAGCTTCGGGAGCTCAGCCCCGAGTGCCTCGCCCGCGCC 306

QY 95 IleArgSerMetValArgGlyTrpGluGlu-----104

DB 307 CTCAGCAGTGTGTAGGCAATACGACGACGAAACCGAGGTGCATCCAAAGGAGGATCC 366

QY 105 -----AlaMetProProLeuGluLys---GlyTrpTrp-----114

DB 367 TTCTACCGGGGGGACCGCCACCGCGCTGCAGCAAGGAGGATGGTGGGGAACCTCTGTA 426

QY 115 -----ProTrpGlyArgGlnGln-----120

DB 427 AAATGGTACTACCAGACCAAACTTCTTCGCAACAGTCATGGCAAGGGCAACAAAGGGTAC 486

QY 121 -----GlnProProGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 133

DB 487 CACCAAGCGTAACCTTCTTCCAGCAGCGGCAAGGGCGCAAGGGTCTTACCAGGT 546

QY 134 TyrTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArg 153

DB 547 TCNACTTCCCG---CAGCAGCAGCAGCAAGGACAA-----CAACAGGACAGAGG 594

QY 154 Gln-----MetTyrPro-----ProCysArgProGlyThrThrGlyGly 166

DB 595 CAGCCATGGTCTTATCAAGTGCAACTTCCCAACACAGCCAGGCAAGGCAAGGG 651

RESULT 2

US-08-706-391B-6

Sequence 6, Application US/08706391B

Patent No. 6174725

GENERAL INFORMATION:

APPLICANT: ANDERSON, OLIN D

TITLE OF INVENTION: ALTERING DOUGH VISCOELASTICITY WITH

MODIFIED GLUTENINS

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: USDA, AGRICULTURAL RESEARCH SERVICE, PACIFIC

WEST

STREET: 800 BUCHANAN STREET

CITY: ALBANY

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94710

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/706,391B

FILING DATE: 30-Aug-1996

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CONNOR, MARGARET A

REGISTRATION NUMBER: 30,043

REFERENCE/DOCKET NUMBER: 0235.95/USDA96-001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 559-6067

TELEFAX: (510) 559-5777

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 477 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-08-706-391B-6

Alignment Scores:

Pred. No.: 0.0166 Length: 477

Score: 130.50 Matches: 55

Percent Similarity: 37.62% Conservative: 53

Best Local Similarity: 27.23% Mismatches: 73

Query Match: 11.37% Indels: 11

DB: 3 Gaps: 11

US-10-053-410-4 (1-206) x US-08-706-391B-6 (1-477)

QY 24 GlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCys-----TrpGlnGlu 40

DB 4 GAAGGTGAAGCGCTCTGAACAG-----CTGCAGTGGCATCGCAACTGCAGAA 51

QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 60

DB 52 CTGCAGGACGCGAACTGAAGCTTGCAGCAGGTTATGGAC-----93

QY 61 GlyGlyGlyGlyValGlyProPheArgTrpGlyLeuArgMetArgCysCys 80

DB 93 -----93

QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100

DB 94 CAGCAGCTGCGTGACATCTCTCCGAATGCCACCGGTTGTTGTTCTCCGGTGTCTGCG 153

QY 101 GlyTyrGluGlu-----AlaMetProProLeuGluLysGlyTyrTrpProTrpGlyArg 118

DB 154 CAGTACGACACACAGATCGTTGTTCCCGCG-----AAAGGTGGTACCTTCTATCCGGGT 207

QY 119 GlnGlnGlnProProGln-----GlyGlyGlyGlyGlnGly 132

DB 208 GAAACCACTCCCGCCGACGAGCTGCAGCAGCGATATCTTCTGGGGCATCCCGGCTCTGCTG 267

QY 133 GlyTyrTyrTyrPro-----CysSerArgProGlyGluGlyTyrGlyTyrGlyGln 149

Db		268	AAGCGTACTACCCGCTGTGTACTTGTCGCCGAG-----	300
Qy		150	GlyGlyGlnArgGlnMetTyPro-----ProCysArgProGlyThrThr---	164
Db		301	-----CAGGTTCTTACTACC CGGGTCAGGCCTTCCGCAGCGTTCTAGTCTTCTTAC	354
Qy		165	-----GlyGlyGlyProArgIleGlyArgValArgLeuThrLySAlaAArgGluTyAla	182
Db		355	CACGTTTCTGTGTAACACACGAGCTGCATCTCTGAAGTTGTCTAAAGCTTCAGACAGCTGGCT	414
Qy		183	AlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePheSerGly	202
Db		415	GCGCAGCTGCCGGCTATGTGCCGCTG-----GAAGGT	447
Qy		203	GlyAsp	204
Db		448	GGTGAC	453

```

RESULT 3
US-09-252-991A-8810/c
; Sequence 8810, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8810
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8810

```

Alignment Scores:	
Pred. No.:	0.0794
Score:	124.00
Percent Similarity:	43.72%
Best Local Similarity:	32.66%
Query Match:	10.80%
DB:	4
Length:	654
Matches:	65
Conservative:	22
Mismatches:	75
Indels:	38
Gaps:	12

US-10-053-410-A (1-206) x US-09-252-991A-8810 (1-654)

```
QY      1 MetAlaLysIleAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20  
       ::::|:::|:::|:::|:::|:::|  
Db     530 CTGGCGCAGCTGAGCTCCGGAGCGGTGCCTTGCTATTGCCGCCCTG----- 480  
  
QY    21 AlaValCysGlnGlyGlu-----ValGluArgGlnArgLeuAspLeuGln 36  
       :::||:::||::||:  
Db   479 -----CGCGCGCATCCCACCCCGCGTTCGCGCGAGCGCGCGCTGCTGGAG 429  
  
QY    37 CysTrpGlnGluValGlnGlusSerProLeuAspAlaCysArgGlnValLeuAspArgGln 56  
       ::::::|:::||||:  
Db   428 GCCTGGGAGAA-----GACGCAGTGTCAATGCGCTGTGTGCGGCCCTGGC-CGATCC 376  
  
QY    57 LeuThrGlyGlyGlyGlyGlyGlyValGlyProPheArg--TpGlyThrGlyLeu 75  
       ||||| |:::|:::|:::|:::|:::|:::|  
Db   375 GGTGCCGCGCGTTGCCGATGCCCGCCGACAGACTTGGCGAACCTCAAGGAACCGCGCCG 316  
  
QY    76 ArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIle 95  
       ||||| |:::|:::|:::|:::|:::|:::|  
Db   315 CGGCGCGCGCTGCTGCC-----GTGGCTCGGGCATGCCGATCCTTCGTCCG 268  
  
QY    96 ArgSerMetValArgGlyTyTrpGluAlaMetProProLeuGluLysGlyTyTrpTPpro 115  
       ||||| |:::|:::|:::|:::|:::|:::|  
Db   267 CGCACAGCTATTGCGTGTGCGCGTAACCTGCGCTGAGAGAGAGCGCGCGCGCGCCCT 208
```

QY		116	TpGlyArgGlnGlnPro-----ProProGlnGlyGlyGlyGlyGln	131
				::::
Dd		207	--GGCGCCCTCGGCATCCGAGCGCGGTGCCGAGCGGTGGCGTCTCGG	151
				::::
QY		132	GlyGlyTyrrTyrrProCysSerArgProGlygluGlyTyrrGlyTyrrGlyelnglGly	151
			::	
Dd		150	----CTGCTCCGCCACCGCGGCCCTGGCCGA-----ACTGCCAGCTGGC	106
QY		152	GlnArgGlnMetTyrrProProCysAraArgProGlyThrThrGlyGlyProArgileGly	171
				::::
Dd		105	CAGCGCGCA-----TGTCGACC CGAGTACGCCGGGGCGCACCGG----	64
QY		172	Argvallarg-LeuthrLysAlaArgGlutyrAlaAlaGlyLeuProMetMetCys	189
				::::
Dd		63	-----CGCCCTCGGCCCTGAGCCGCGAG---GCCAGCGTGTGCCGGCGCTCTGC	18
				::::

```

RESULT 4
US-09-252-991A-9241
; Sequence 9241, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9241
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9241

```

Alignment Scores:	
Pred. No.:	0.168
Score:	124.00
Percent Similarity:	43.72%
Best Local Similarity:	32.66%
Query Match:	10.80%
DB:	4
Gaps:	12
Indels:	38
Mismatches:	75
Conservative:	22
Matches:	65
Length:	1233

US-10-053-410-4 (1-206) x US-09-252-991A-9241 (1-1733)

Qy	1	MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal	20
Db	86	CTGCCGACCTGGAGCTCCGGAGCGCTGCCGTGTGATTGGCGCCCTG	136
Qy	21	AlaValCysGlnGlyGlu-----ValGluArgGlnArgLeuArgAspLeuGln	36
Db	137	-----CGCGCGATCCGACCCCGCGCTGCGCGAGCGCGCGCTGCTGGAG	187
Qy	37	CysTrpClnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGln	56
Db	188	GCCTGGAGGAA-----GACCGAGTGGTCGATCGCTCTGTGGCGCCCTGGC-CAATCC	240
Qy	57	LeuThrGlyGlyGlyGlyGlyValGlyPropheArg---TrpGlyThrGlyLeu	75
Db	241	GGTGCCGGCGGTGGCCGATGCCGCCGCGCAGAGCTTGGCGAACTCAGGAACCGCGCCG	300
Qy	76	ArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIle	95
Db	301	CGGGCGCGCTGCTGCC-----GTGGCTCGGCATGCCGATGCCTTCTGCTCGG	348
Qy	96	ArgSerMetValArgGlyTyrGluGluAlaMetProProLeuGluLysGlyTrpTrpPro	115
Db	349	CGCCAGCGTATTGGCTGGCTGGCGAACTCGGCTGGAGAGAGCGCGCGCGCCCT	408

```
QY 116 TrpGlyArgGlnGlnGlnPro-----ProProGlnGlyGlyGlyGlyGln 131
Db 409 ---GGCGCCCTCGGCATCCGACGCGCGGTGCGCGGAGCGGTGCGGTCTCGG 465
QY 132 GlyGlyTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlnGlyGly 151
Db 466 -----CTGGCTCGGCACACGCGGCTTGGCCGA-----ACTGGCCAGGCTGGC 510
QY 152 GlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyGlyProArgIleGly 171
Db 511 CAGCGCCGA-----TGTGACCCGAGGTACGCGCGCGCCACCGG-----552
QY 172 ArgValArg-LeuThrLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCys 189
Db 553 -----CGCCCTCGGCCTGAGCGCGAG---GGACGGTGTCTGCCGCGCTCTGC 598

RESULT 5
US-09-252-991A-9071
; Sequence 9071, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9071
; LENGTH: 1848
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9071

Alignment Scores:
Pred. No.: 0.271 Length: 1848
Score: 124.00 Matches: 65
Percent Similarity: 43.72% Conservative: 22
Best Local Similarity: 32.66% Mismatches: 75
Query Match: 10.80% Indels: 38
DB: 4 Gaps: 12

US-10-053-410-4 (1-206) x US-09-252-991A-9071 (1-1848)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
Db 979 CTGCGCCGACCTGGAGCTGCGGAGCGCTGCGGTGCTGATTGCGGCCCTG-----1029
QY 21 AlaValCysGlnGlyGlu-----ValGluArgGlnArgLeuArgAspLeuGln 36
Db 1030 -----CGCGCGATCCGACCCCGCGGTGCGCGCGAGCGCGCGCTGCTGGAG 1080
QY 37 CysTrpGlnGluValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 56
Db 1081 GCCTGGAGGAA-----GACGCAAGTGTGATGCGCTGTGCGCGCCCTGGC-CGATCC 1133
QY 57 LeuThrGlyGlyGlyGlyGlyGlyValGlyValGlyProPheArg---TrpGlyThrGlyLeu 75
Db 1134 GTGCGCGCGGTGCGCGCGCGCGAGAGCTTGGCGAACTCAAGGAACCGCGCGC 1193
QY 76 ArgMetArgCysCysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 95
Db 1194 CGGCGCGCGCTGCTGCC-----GTGGCTCGGCGATCGCGATCGCTTCGTCG 1241
QY 96 ArgSerMetValArgGlyTyrGluGluAlaMetProLeuGluLysGlyTyrTrpPro 115
Db 1242 CGCCAGCGTATTGCGGTGCGCGAACTGCGCGTGGAGGAGAGCGCGCGCGCCCT 1301
QY 116 TrpGlyArgGlnGlnGlnPro-----ProProGlnGlyGlyGlyGlyGln 131
```

```
Db 1302 ---GGCGCCCTCGGCATCCGACGCGCGGTGCGCGGAGCGGTGCGGTCTCGG 1358
QY 132 GlyGlyTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlnGlyGly 151
Db 1359 -----CTGGCTCGGCACACGCGGCTTGGCCGA-----ACTGGCCAGGCTGGC 1403
QY 152 GlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyGlyProArgIleGly 171
Db 1404 CAGCGCCGA-----TGTGACCCGAGGTACGCGCGCGCCACCGG-----1445
QY 172 ArgValArg-LeuThrLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCys 189
Db 1446 -----CGCCCTCGGCCTGAGCGCGAG---GGACGGTGTCTGCCGCGCTCTGC 1491

RESULT 6
US-09-252-991A-8712/c
; Sequence 8712, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8712
; LENGTH: 1980
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8712

Alignment Scores:
Pred. No.: 0.294 Length: 1980
Score: 124.00 Matches: 65
Percent Similarity: 43.72% Conservative: 22
Best Local Similarity: 32.66% Mismatches: 75
Query Match: 10.80% Indels: 38
DB: 4 Gaps: 12

US-10-053-410-4 (1-206) x US-09-252-991A-8712 (1-1980)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
Db 942 CTGCGCCGACCTGGAGCTGCGGAGCGCTGCGGTGCTGATTGCGGCCCTG-----892
QY 21 AlaValCysGlnGlyGlu-----ValGluArgGlnArgLeuArgAspLeuGln 36
Db 891 -----CGCGCGATCCGACCCCGCGGTGCGCGCGAGCGCGCGCTGCTGGAG 841
QY 37 CysTrpGlnGluValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 56
Db 840 GCCTGGAGGAA-----GACGCAAGTGTGATGCGCTGTGCGCGCCCTGGC-CGATCC 788
QY 57 LeuThrGlyGlyGlyGlyGlyGlyValGlyValGlyProPheArg---TrpGlyThrGlyLeu 75
Db 787 GTGCGCGCGGTGCGCGCGCGCGAGAGCTTGGCGAACTCAAGGAACCGCGCGC 728
QY 76 ArgMetArgCysCysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 95
Db 727 CGGCGCGCGCTGCTGCC-----GTGGCTCGGCGATCGCGATCGCTTCGTCG 680
QY 96 ArgSerMetValArgGlyTyrGluGluAlaMetProLeuGluLysGlyTyrTrpPro 115
Db 679 CGCCAGCGTATTGCGGTGCGCGAACTGCGCGTGGAGGAGAGCGCGCGCGCCCT 620
QY 116 TrpGlyArgGlnGlnGlnPro-----ProProGlnGlyGlyGlyGlyGln 131
```


Db 619 -----GGCCGCCCTCGGCATCCGACGCGCGGTGCGCCCGAGCGGTGCGGTGCTCGG 563
QY 132 GlyGlyTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGly 151
Db 562 -----CTGGCTCCGCGCACCGCGCCCTGGCCGA-----ACTGCGCAGGCTGCG 518
QY 152 GlnArgGlnMetTyrProProCysArgProGlyThrGlyGlyProArgGlyGly 171
Db 517 CAGCGCCGA-----TGTGACCGGAGGTACGCGCGCGCGCCACCGG----- 476
QY 172 ArgValArg-LeuThrLysAlaArgGlyTyrAlaAlaGlyLeuProMetMetCys 189
Db 475 -----CGCCCTCGGCTGAGCGCGAG---GCGACGGTGTGCGCGCGCTGCTGC 430

RESULT 7

US-09-010-928B-3
; Sequence 3, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph P
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREOF
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28977
REFERENCE/DOCKET NUMBER: 1447-109P
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2824 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: -
LOCATION: 1..2824
OTHER INFORMATION: /note= "flagelliform DNA sequence taken from 3' region. Stop codon begins at position 2722."
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2724
US-09-010-928B-3

Alignment Scores:

Pred. No.:	0.537	Length:	2824
Score:	123.00	Matches:	48
Percent Similarity:	38.93%	Conservative:	10
Best Local Similarity:	32.21%	Mismatches:	64
Query Match:	10.71%	Indels:	27
DB:	2	Gaps:	5

US-10-053-410-4 (1-206) x US-09-010-928B-3 (1-2824)

QY 32 LeuArgAspLeuGlnCysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGln 51

Db 715 ATAGAAGACTTGATATTAATTCATGCGCTGATGCGCCGATA----- 759
QY 52 ValLeuAspArgGlnLeuThr-----GlyGlyGlyGlyGlyGlyValGly 67
Db 760 ACGATTTCAGAAAGTAATTAACAAATTAGTGTGTGGAGGTTCGCGACCCCGTGTCTGGA 819
QY 68 ProPheArgTrpGlyThrGly--LeuArgMetArgCysCysGlnGlnLeuGlnAspValS 87
Db 820 CCAGGAGGTGTAGGCGCTGTGTGCTCCGACCCAGAGGTGTAGGACCTGGAGGTTCTGGA 879
QY 87 erArgGluCysArgCysAlaAlaAlaLeuArgSerMetValArgGlyTyrGluGluAlaMetP 107
Db 880 CCAGGAGGTGTAGGACCTGTGTCTGCTGCTCCAGGAGCGT----- 920
QY 107 roProLeuGlnLysGlyTrpTrpProTrp-GlyArgGln-----GlnGlnPro 122
Db 921 -----AGGACCTGGTGTCTGCTGACCTTACGGACCTGGCGGTTCTGGACCT 969
QY 123 ProProGlnGlyGlyGlyGlnGlyGlyTyrTyrTyrProCysSerArgProGly 142
Db 970 GGAGGTGCGAGCGAGCTGGAGGACCTGTGTGGACATACGGACCTGGTGGATCATATGA 1029
QY 143 GluGlyTyrGlyTyrGlyGlnGlyGlnArgGlnMetTyrProProCysArgProGly 162
Db 1030 CCTGGTGTTCGCGAGGACCCGCTGTGTGCTGCGCGACCATACGGACCTGGAGGTGAAGGA 1089
QY 163 ThrThrGly--GlyGlyPro 168
Db 1090 CCCGCTGGTGTGCGGACCC 1110

RESULT 8

US-08-899-595-4/c
; Sequence 4, Application US/08899595
; Patent No. 6111072
; GENERAL INFORMATION:
; APPLICANT: Narumiya, Shuh
; APPLICANT: Takahashi, No. 6111072uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
; TITLE OF INVENTION: ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,595
FILING DATE: 24-JUL-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-242701
FILING DATE: 26-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 9-90170
FILING DATE: 25-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Stephen A. Bent
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 049441/0112
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

```

; ; LENGTH: 5822 base pairs
; ; TYPE: nucleic acid
; ; STRANDEDNESS: single
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: cDNA to mRNA
; ; ORIGINAL SOURCE:
; ; ORGANISM: Human
; ; FEATURE:
; ; NAME/KEY: CDS
; ; LOCATION: 28..3972
; ; US-08-899-595-4

```

Alignment Scores:	
Pred. No.:	2.88
Score:	118.50
Percent Similarity:	43.09%
Best Local Similarity:	34.25%
Query Match:	10.32%
DB:	3
Length:	5822
Matches:	62
Conservative:	16
Mismatches:	63
Indels:	42
Gaps:	10

US-10-053-410-4 (1-206) X US-08-899-595-4 (1-5822)

Qy	18	ValAlaValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCys	37
Db	2383	GTGAGGTGGAGGAATGCCAGGGCCTCCGGGAAATGAGGAGGTGGAGGATTCACGAC	2324
Qy	38	TrpGlnGluValGlnGlnSerPro-----LeuAspAlaCysArgGlnVal	52
Db	2323	CACCAGAAAGAGGGGAGGAGGAGTGGCAATTCTGCTTCTCCAGGCACAGGAGGAGGTG	2364
Qy	53	-----LeuAspArgGlnLeuThrGlyGlyGlyGlyGly-----	64
Db	2263	GGGGGGGAATTCAGCACATCCAGGC-AAAGGAGGTGGTGGTGGGGGATTCCTAGCACCTC	2205
Qy	65	---GlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCysGlnGlnLeu	83
Db	2204	CCAGGCAAAAGGAGG---AGTGGGGGATGGCAGTACCTCCAGGCCAAAGAGGGGTGA	2149
Qy	84	GlnAspValSer-----ArgGluCysArgCysAlaAlaIleArgSerMet	98
Db	2148	AGGGATGCCACACCTCTCAGGCCAAAGAGGGGTGGAGGGATGGTACATCCCGACAA	2089
Qy	99	ValArgGlyTyr-GluGluAlaMetProProLeuGluLysGlyTrpTrpProTrpGlyAr	118
Db	2088	AGGAGGGGTGGAGAGATAGCAGTACCTCCAGTAAAGAGG-----GGTGA	2041
Qy	118	gGlnGlnProProPro-----GlnGlyGlyGlyGlyGlnGlyGlyTyrTyrPr	137
Db	2040	GGAGATGCAAAACACCCCGAGCAAAAGAGGTGGAGGAGGAGGAGGAGGA---	1992
Qy	137	oCysSerArgProGlyGluGlyTyrGlyTyr-----GlyGlnGlyGlyGlnAr	153
Db	1991	-----GGAGGAGGAGGAGGAGTGGTACTATCCCGAGGACAGGTGGTGGTGG	1945
Qy	153	gGlnMetTyrPro-----ProCysArgProGlyThrThrGlyGlyGlyProArgIleG	171
Db	1944	AATAATAGTCCAGAGTCACCAAGTAAAGAGGGGGGACAGGAGCAGCACTAGG	1885
Qy	171	Y 171	
Db	1884	A 1884	

RESULT 9

```

US-08-899-595-5
; Sequence 5, Application US/08899595
; Patent No. 6111072
; GENERAL INFORMATION:
; APPLICANT: Natumiya, Shuh
; APPLICANT: Takahashi No. 6111072uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
; TITLE OF INVENTION: ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:

```



```
QY      64 GlyGlyValGly---ProPheArgTrpGlyThrGlyLeuArgMetArgCysCysGlnGln 82
Db      2476 GGAGGGGGAGCGCACCCCGAGAGGGCGGACCGAGGGCTTCGGTCTCGTCAGGGGCCCCGG 2417

QY      83 LeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArgGlyTyr 102
Db      2416 GAAGGAGGCTGGGGTCCCCGGGGCGTGCCCC-----GGAGGGGGGAAGGCTTT 2366

QY     103 GluGluAlaMetProProLeuGluLys-----GlyTrpTrp 114
Db     2365 CCAGGAGCCCGACCTCCAGGGCGCGAAGAGGTCTCTGGGGCGGCCCGGGGGAGGCTTC 2306

QY     115 ProTrpGlyArg-----GlnGlnGlnProProGlnGlyGly----- 127
Db     2305 CCGTGAGGGCGAGGACCTTCTGGAGGCGCCCGCTCCAGAAAGAGGCCCTTCGGGGCC 2246

QY     128 -----GlyGlyGlyGlnGlyGlyTyrTyrTyrProCysSerArgProGlyGlu 143
Db     2245 GGAGAACCCAGGGGAGGTCCAGGGGGGAGGACCCGCTCCGTGCGGT---CCAGGCCCC 2189

QY     144 GlyTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProProCysArgProGlyThr 163
Db     2188 CGTCTTGGACGAGCGGAGCTGGAGCCAGACCCCTCCCGCGCGAGTCCCG----- 2135

QY     164 ThrGlyGlyGlyProArgIleGlyArgValArgLeuThrLysAlaArgGluTyrAlaAla 183
Db     2134 GCCGGAGGAGTCCACGG-----CCTCGGGCACGGGACCCCGGGTGGC 2093

QY     184 Gly 184
Db     2092 GGC 2090
```

Search completed: November 29, 2003, 19:16:29
Job time : 1629 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2003, 13:23:43 ; Search time 317 Seconds
(without alignments)
8089.799 Million cell updates/sec

Title: US-10-053-410-3
Perfect score: 950
Sequence: 1 aaaaaaacccctcgat.....aaaaaaaaaaaaaaaaaaaaa 950

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91.2	9.6	2296	19 AAV20662	Barley D-hordein g
2	91.2	9.6	2380	22 AAF63380	Barley cDNA involv
3	91.2	9.6	2434	22 AAF63379	Barley cDNA involv
4	72.4	7.6	114955	20 AAX53491	Human adenosine A1
C 5	64.6	6.8	1614	22 AAH26499	Human low density
C 6	64.6	6.8	12425	22 AAH26495	Human low density
C 7	63.4	6.7	3957	22 AAA09686	HSV-2 immediate ea
8	63.4	6.7	154746	24 AAD25519	Human herpesvirus

C	9	63.4	6.7	154746	24	AD25519	Human herpesvirus
10	63	6.6	88421	24	AA40781	88421nt genomic DN	
11	62.4	6.6	1272	20	AA09010	Brn-3a polynucleot	
12	62.4	6.6	1272	21	AA29006	Human transcriptio	
C	13	61.6	6.5	114955	20	AAX53491	Human adenosine A1
14	60	6.3	1127	21	AA02477	Human colon cancer	
15	59.4	6.3	4524	20	AAV33912	Nucleotide sequenc	
C	16	59	6.2	3035	24	ABK09774	Human ovarian tumo
C	17	59	6.2	3056	24	ABK84722	Human cDNA differe
C	18	59	6.2	3056	25	ACA56555	Human signalling p
C	19	59	6.2	6695	22	AAK70340	Human immune/haema
C	20	59	6.2	6695	22	AAK82282	Human immune/haema
21	57.8	6.1	43411	24	ABQ88169	Human osteoblast d	
C	22	57.4	6.0	3169	22	AAI58822	Human polynucleoti
C	23	57.4	6.0	3203	22	AAI60608	Human polynucleoti
C	24	57.4	6.0	4190	15	AAQ55350	Sequence of the un
C	25	56.4	5.9	1556	18	AA794535	Maize cinnamoyl Co
26	56	5.9	2549	25	ABZ24709	Human cell growth,	
27	56	5.9	2810	25	ABZ24708	Human cell growth,	
C	28	56	5.9	3147	23	AA85126	DNA encoding novel
C	29	56	5.9	65140	22	AAI17184	Streptomyces nous
C	30	56	5.9	125401	22	AAI17186	Streptomyces nous
C	31	55.8	5.9	6225	20	AA55273	Human enzyme-relat
C	32	55.8	5.9	6225	21	AA20843	Human multiple tar
C	33	55.8	5.9	6225	21	AA34721	Human adenosine re
C	34	55.8	5.9	35384	21	AA21436	Human enzyme-relat
C	35	55.6	5.9	594	24	ABQ43958	Oligonucleotide fo
C	36	55.6	5.9	594	24	ABQ43959	Oligonucleotide fo
C	37	55.6	5.9	5811	24	AB578655	S. macromyceticus
38	55.4	5.8	1000	21	AAA02484	Human colon cancer	
39	55.4	5.8	1698	23	AA84408	DNA encoding novel	
40	55.4	5.8	2049	23	AA876210	DNA encoding novel	
C	41	55.4	5.8	2561	22	AAH26500	Rabbit low density
C	42	55.4	5.8	4403765	22	AAI99683	Mycobacterium tube
C	43	55.2	5.8	1511	24	ABT11009	Human breast cance
C	44	55.2	5.8	2668	22	AAK75409	Human immune/haema
C	45	55	5.8	2687	21	AAZ30700	S.clavuligerus cla

ALIGNMENTS

RESULT 1
AAV20662
ID AAV20662 standard; DNA; 2296 BP.
XX AAV20662;
AC AAV20662;
XX
DT 23-JUN-1998 (first entry)
XX
DE Barley D-hordein gene 5'-terminal region.
XX
KW Barley; D-hordein; DPP3; promoter; gene expression; regulation;
KW seed; structural gene; ds.
XX
OS Hordeum vulgare.
XX
PN WO9803655-A2.
XX
PD 29-JAN-1998.
XX
PF 22-JUL-1997; 97WO-JP02526.
XX
PR 23-JUL-1996; 96JP-0193433.
XX
PA (SAPP) SAPPORO BREWERIES.
PI Hirota N, Ito K, Kihara M, Kuroda H;
XX
DR WPI; 1998-120779/11.
XX
PT Gene expression regulatory DNA, expression cassettes and vectors -
comprising promoter region from barley, Hordeum vulgare, D-hordein

PT gene, useful to control expression of desired gene e.g. to improve
XX seeds
PS Example 3; Page 25-26; 42pp; English.
XX
CC The present sequence represents the 5'-terminal region of the barley
CC D-hordein gene. The present invention describes gene expression
CC regulatory DNA which comprises a promoter region derived from the
CC barley (Hordeum vulgare) D-hordein gene which allows expression of a
CC desired structural gene, and a regulatory region regulating such
CC expression. The introduction into plants of expression cassettes
CC containing the gene expression regulatory DNA (either directly or via
CC expression vectors) enables the expression of a gene within a plant cell
CC e.g. barley to be controlled. The use of activating and suppressing
CC regions in the regulatory DNA allows control of expression by e.g.
CC tissue type or developmental stage, whilst the use of only an activating
CC region maintains expression at a high level, providing an effective
CC production means when recovery of the product of the structural gene is
CC desired. The expression cassette/vector may be introduced into e.g.
CC maturing seed endosperm tissue or regeneratable plant cells (e.g. derived
CC from anthers) to improve seeds of barley or other plants, to produce
CC gene products in seeds or to contribute to plant breeding programmes.
CC The expression regulatory DNA can also be used in expression systems in
CC vitro. GUS activity in barley protoplasts transfected with plasmid
CC DHP3GUS2 comprising isolated D-hordein promoter region was 1.5 times
CC higher than in those transfected with control pACTIF.
XX
SQ Sequence 2296 BP; 629 A; 708 C; 606 G; 353 T; 0 other;

Query Match 9.6%; Score 91.2; DB 19; Length 2296;
Best Local Similarity 61.4%; Pred. No. 2.6e-07;
Matches 178; Conservative 0; Mismatches 88; Indels 24; Gaps 1;

QY 207 AGGACCTCGAGTGTGGCAGAGGTCTCCAGAGAGCCCGCTGACCGGTGCGCCAGGTGTC 266
DB 145 AGGCAGCTACAGTGTGAGCGCGAGCTCCAGAGAGTCTGCTGAGGCGTGC CGCGGGTTC 204
QY 267 CTCGACCGGACGCTAACCGGCGCGCGCGGGTGGCCCGTTCGCGTGGGGC 326
DB 205 GTGGACCAACAGCTG-----GTTGGCCAGCTGCATGGAGC 240
QY 327 ACCGGCTCCGATGCGGTGCTGCCAGCAGCTCCAGGAGCTGAGCCGAGTGC CGCTGC 386
DB 241 ACGGGGCTCCAGTGCAGTGTGCTGCCAGCAGCTTCGGGAGCTCAGCCCGAGTGC CGCCCC 300
QY 387 GCCGCATCCGAGCATGTGTACGGGCTACGAGAGGCCATGCCCGCTGGAGAAAGGC 446
DB 301 GTCGCCCTCAGCCAGCTGCTGAGGCAATACGAGCAGCAAAACGAGGTGCCATCCAGGGA 360
QY 447 TGGTGGCCATGGGGCGGCGAGCAGCAGCCCGCCCGCAGGAGGAGG 496
DB 361 GGATCCTTCTACCGGGCGGGACCGCACC CGCGCTGCAGCAAGGAGGATG 410

RESULT 2
AAF63380
ID AAF63380 standard; cDNA; 2380 BP.
XX
AC AAF63380;
XX
DT 10-MAY-2001 (first entry)
XX
DE Barley cDNA involved in D-hordein production SEQ ID 2.
XX
KW Barley; gel protein; D-hordein; malt; wort filterability; extraction; ss.
XX
OS Hordeum vulgare.
XX
PN WO200111946-A1.
XX
PD 22-FEB-2001.
XX
PF 16-AUG-2000; 2000WO-JP05476.
XX

XX 16-AUG-1999; 99JP-0229696.
XX (SAPB) SAPPORO BREWERIES LTD.
XX
PI Hirota N, Kihara M, Ito K;
XX
DR WPI; 2001-191587/19.
XX
PT Transformation of barley with a D-hordein expression regulator for
PT production of barley with lowered gel protein content -
XX
PS Claim 4; Page 33-34; 40pp; Japanese.
XX
CC This invention relates to barley having a low gel protein content. The
CC barley is transformed with a polynucleotide sequence which regulated the
CC formation of D-hordein. Transformation results in the production of
CC barley strains with improved malting properties such as wort
CC filterability and efficiency of extraction. The present sequence
CC represents cDNA involved in the regulation of D-hordein production.
XX
SQ Sequence 2380 BP; 642 A; 747 C; 622 G; 369 T; 0 other;

Query Match 9.6%; Score 91.2; DB 22; Length 2380;
Best Local Similarity 61.4%; Pred. No. 2.6e-07;
Matches 178; Conservative 0; Mismatches 88; Indels 24; Gaps 1;

QY 207 AGGACCTCGAGTGTGGCAGAGGTCTCCAGAGAGCCCGCTGACCGGTGCGCCAGGTGTC 266
DB 132 AGGCAGCTACAGTGTGAGCGCGAGCTCCAGGAGAGTCTGCTGAGGCGTGC CGCGGGTTC 191
QY 267 CTCGACCGGACGCTAACCGGCGCGCGGGTGGCCCGTTCGCGTGGGGC 326
DB 192 GTGGACCAACAGCTG-----GTTGGCCAGCTGCATGGAGC 227
QY 327 ACCGGCTCCGATGCGGTGCTGCCAGCAGCTCCAGGAGCTGAGCCGAGTGC CGCTGC 386
DB 228 ACGGGGCTCCAGTGCAGTGTGCTGCCAGCAGCTTCGGGAGCTCAGCCCGAGTGC CGCCCC 287
QY 387 GCCGCATCCGAGCATGTGTACGGGCTACGAGGAGGCCATGCCCGCTGGAGAAAGGC 446
DB 288 GTCGCCCTCAGCCAGTGTGAGGCAATACGAGCAGCAAAACGAGGTGCCATCCAGGGA 347
QY 447 TGGTGGCCATGGGGCGGCGAGCAGCAGCCCGCCCGCAGGAGGAGGAGG 496
DB 348 GGATCCTTCTACCGGGCGGGACCGCACC CGCGCTGCAGCAAGGAGGATG 397

RESULT 3
AAF63379
ID AAF63379 standard; cDNA; 2434 BP.
XX
AC AAF63379;
XX
DT 10-MAY-2001 (first entry)
XX
DE Barley cDNA involved in D-hordein production SEQ ID 1.
XX
KW Barley; gel protein; D-hordein; malt; wort filterability; extraction; ss.
XX
OS Hordeum vulgare.
XX
PN WO200111946-A1.
XX
PD 22-FEB-2001.
XX
PF 16-AUG-2000; 2000WO-JP05476.
XX
PR 16-AUG-1999; 99JP-0229696.
XX
PA (SAPB) SAPPORO BREWERIES LTD.
XX
PI Hirota N, Kihara M, Ito K;

Db 108854 CGCCTGGCTCGCCTGGCGCCGCCNNNNNGGGGGCGGCGGCGCTGGCTGGCTCGCCTBGGGC 108913
QY 662 GTTGGCGATGATGCGCGGTGTCGAGAGCCCGAGGAGTGAGCATCTTCTCCGGCGCGCA 721
Db 108914 CCSNNNDNNGGGCGCGCGCTGGCTGGCTBGGGCCCCNNNDNNGGGCGCGCGC 108973
QY 722 C 722
Db 108974 C 108974

RESULT 5
AAH26499/c
ID AAH26499 standard; DNA; 1614 BP.
XX
AC AAH26499;
XX
DT 12-NOV-2001 (first entry)
XX
DE Human low density lipoprotein binding protein 2 (LBP-2) DNA.
XX
KW Low density lipoprotein binding protein 2; LBP-2; LDL; human;
KW atherosclerosis; antiarteriosclerotic; gene therapy; diagnosis;
KW vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200164874-A2.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06356.
XX
PR 02-MAR-2000; 2000US-0517849.
PR 14-JUL-2000; 2000US-0616289.
XX
PA (BOST-) BOSTON HEART FOUND INC.
XX
PI Lees AM, Lees RS, Law SW, Arjona AA;
XX
XX WPI; 2001-565505/63.
DR P-PSDB; AAB82806.
XX
PT New isolated low density lipoprotein binding polypeptide for treating,
PT diagnosing and/or identifying therapeutic agents for atherosclerosis -
XX
PS Claim 7; Fig 7A; 143bp; English.

XX The present sequence is that of the coding region of the human
CC gene (see also AAH26494) encoding novel human low density lipoprotein
CC binding protein 2 (LBP-2, see AAB82806). The gene was isolated
CC from a genomic DNA library using LBP-2 cDNA as probe. The LBP-2
CC protein predicted from the present sequence differs from that
CC predicted from a cDNA clone (see AAB82803) in that it contains an
CC additional 321 amino acids at its N-terminus (the cDNA is a 5'
CC truncation). LBP-2 nucleic acids are among claimed polynucleotides
CC of the invention that encode novel polypeptides capable of binding
CC to native and methylated LDL. Also claimed are isolated LBP
CC polypeptides, and biologically active fragments and analogues of
CC them, as well as expression vectors, cells and methods of producing
CC the LBPs. Methods of determining if an animal is at risk for
CC atherosclerosis, methods for evaluating an agent for use in treating
CC atherosclerosis, and methods for treating a cell having an
CC abnormality in structure or metabolism of LBP are claimed.
CC Pharmaceutical compositions comprising an LBP polypeptide or nucleic
CC acid, and vaccine compositions, are also claimed.
XX
SQ Sequence 1614 BP; 243 A; 638 C; 564 G; 169 T; 0 other;

Query Match 6.8%; Score 64.6; DB 22; Length 1614;
Best Local Similarity 46.4%; Pred. No. 0.012;
Matches 211; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 108 GAGATGGCTAAGATCGCCCGCGCGCGCGCGCGCTGTCTTCGCGCCCTGTGGTGC 167
Db 774 GAGGTAGCGCAGCATTTCCCGCAGGCTCAGCGGCGCGCGCGCGCCACCGC 715
QY 168 GTGGCGGTCTGCAAGCGAGGTGAGCGGCGAGAGGCTCAGGACCTGAGTGTCTGGCAG 227
Db 714 GCGCCCTCTCCGCGCGCGCTGTGGTGTGCGCGCGCGCGCGCTGTCTGTCTGTGGCGCGC 655
QY 228 GAGGTCCAGGAGAGCCCGCTCGACGGGTGCGCCCGAGTCTCTGACCGGCGAGCTAACCGGC 287
Db 654 CGCGCGCTGTGCGCGCGCGCGCAGCGCGCGCTCCCGGGCGCGACGCGCGGGGGGGG 595
QY 288 GCGCGCGCGCGCGCGCGCTTGGCCCGTTCCGTTGGGCGACCGGGCTCCGGATGCCGTGC 347
Db 594 GCG 535
QY 348 TGCCAGAGCTCCAGGACGTGAGCGCGAGTCCCGTGC CGCGCCCATTCGAGAGCATGGTC 407
Db 534 GGCAGGGGCGCGCGCGCTGCGCGCGGGGCGCGCGCTCGCGGGGCGCGCGCGCGCGCGG 475
QY 408 AGGGGCTACGAGAGGCGCATGCGCGCGCTGGAGAGGCTGGTGGCCATGGGGCGCGCAG 467
Db 474 CGCTGTGCG 415
QY 468 CAGCAGCG 527
Db 414 CGCGCGCGTGTGCG 355
QY 528 CGCGCAGGAGAGGATACCGCTACCGTACCGTACCGTACCGTACCGTACCGTACCGTACCG 562
Db 354 GCG 320

RESULT 6
AAH26495/c
ID AAH26495 standard; DNA; 12425 BP.
XX
AC AAH26495;
XX
DT 12-NOV-2001 (first entry)
XX
DE Human low density lipoprotein binding protein 2 (LBP-2) gene.
XX
KW Low density lipoprotein binding protein 2; LBP-2; LDL; human;
KW atherosclerosis; antiarteriosclerotic; gene therapy; diagnosis;
KW vaccine; ds.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT CDS 2832..5153
FT /tag= a
FT /note= "includes introns"
FT exon 2832..3785
FT /tag= b
FT intron 3786..4207
FT /tag= c
FT exon 4208..4502
FT /tag= d
FT intron 4503..4593
FT /tag= e
FT exon 4594..4694
FT /tag= f
FT intron 4695..4787
FT /tag= g
FT exon 4788..4899
FT /tag= h
FT intron 4900..4994
FT /tag= i
FT exon 4995..5153
FT /tag= j
XX
PN WO200164874-A2.

QY 428 GCCGCGCTGGAGAAAGCTGGTGGCCATCGGGGCGGCACGACGACCGCGCGCGAGG 487
Db 2287 TCTGGAGAGCAGGTCCGCGCGCGCGCGCGGAGCTACAGCGCGCGGCTCGCG 2228
QY 488 AGGAGGAGGAGGACAGGGGGCTACTACTACCCCTCGAGCCGCCAGGAGGAGATACGG 547
Db 2227 GCAGCGCGGCGCCAGGCGCGCGCGCGACGAGGCTACGCGCGCGACCGCGCGG 2168
QY 548 CTACGGTCAGGTGGCAGCGCGGAGATGATTCACCGTGTCTCCCGGACACCGCGG 607
Db 2167 CTTCTGCTGCGCGCGCGCGCGAGGT-----CCCGCGCAGGCGCATCAGCACCGCGC 2113
QY 608 CGGGCCAAAGGATCGCGCGCTGAGGCTTACGAAGGCCCGGAGTACGCGCGGGTTGCC 667
Db 2112 GTCGCGCAGACCGAGCTCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGG 2053
QY 668 GATGATGTCCCGCTGTGCGAGCCC 692
Db 2052 CGGGCG 2028

RESULT 8
ID AAD25519 standard; DNA; 154746 BP.

AC AAD25519;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Human herpesvirus 2 complete DNA genome.

DE Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
KW antibacterial; fungicide; protozoicide; antirheumatic; antiinflammatory;
KW antiarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW vasculitis; ds.

XX
OS Human herpesvirus 2.
XX
PN WO200176643-A1.

XX 18-OCT-2001.

XX 06-APR-2001; 2001WO-US11372.

XX 07-APR-2000; 2000US-195680P.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Orson FM, Kinsey BM, Bhogal BS;

XX WPI; 2002-066308/09.

XX Composition for oral delivery of vaccines, comprises expression vector
PT containing antigenic genomic sequence, bound to aggregated
PT protein-polycationic polymer conjugate or suspension

PS Disclosure; Page 90-132; 145pp; English.

XX The invention relates to a composition comprising an expression vector
CC bound to an aggregated protein-polycationic polymer conjugate or
CC suspension. The expression vector contains a promoter polynucleotide
CC sequence operatively linked to a polynucleotide sequence encoding an
CC antigen which is a fragment of a gene or genome associated with an
CC infectious disease, cancer and autoimmune disease such as rheumatoid
CC arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
CC consisting of bacterium, fungus, protozoa and virus such as human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
CC virus (HCV), influenza and respiratory syncytial virus (RSV), and
CC optionally comprising a nucleotide sequence encoding a cytokine (or a
CC cytokine expression vector), is useful for inducing an immune response
CC (systemic and/or mucosal) in an organism. The cytokine expression vector

CC contains a sequence for granulocyte macrophage-colony stimulating factor
CC (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
CC the antigen and the cytokine are under transcriptional control of same or
CC different promoter polynucleotide sequences. The expression vector, as a
CC DNA vaccine is useful for treating a condition in an organism. The
CC present sequence is human herpesvirus 2 complete DNA genome related
CC to the invention.

XX Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 other;

QY Query Match 6.7%; Score 63.4; DB 24; Length 154746;
Best Local Similarity 45.8%; Pred. No. 0.027;
Matches 259; Conservative 0; Mismatches 301; Indels 5; Gaps 1;

QY 128 GCGCGCGCGCGCGCGCTGCTTCGCGCCCTGTGCGCGCTGCGCGCTGCGCGCGCGA 187
Db 129446 GCGCGCGCGCGCGCGCTGCGCGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCG 129505
QY 188 GGTGAGCGCGCAGAGGCTCAGGAGCTGTCAGTGTCTGGCAGGAGTCCAGGAGAGCCCGCT 247
Db 129506 GCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 129565
QY 248 CGACGGTCCCGCAGGTCTCTCGACGCGAGTAAACGCGCGCGCGCGCGCGCGCGCGT 307
Db 129566 CGCGCGCGCGCGCGCGCGCGCTCCGCGCGCTCTTCTTCGCGCGCGCGCGCGCG 129625
QY 308 TGGCCCGTTCGCGTGGGCGCACCGGCTCGGATCGGTGCTCCAGCAGCTCCAGGACGT 367
Db 129626 GCGCGCGCGCTGCGCGCGCGCGCGCTTTCGCGCTTCGCGCGCTCCCGCGCGCG 129685
QY 368 GAGCGCGAGTCCGCTGCGCGCGCATCCGAGCATGTGTACGGGCTACGAGGAGGCCAT 427
Db 129686 GCGCGCGCGCAGTTCGCGCGCGCGCGCGTGTCTCGCGCAGCAGGGCGCGCGCTCTGT 129745
QY 428 GCGCGCGCTGGAGAAAGGTGTGTGCGCATGTGGGCGCGCAGCAGCAGCGCGCGCGAGG 487
Db 129746 TCTGGAAGAGCAGGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCGCG 129805
QY 488 AGGAGGAGGAGCAGGGGGGTACTACTACCCCTGACAGCCCGCAGGAGGAGGATACGG 547
Db 129806 GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCACGCGCGCGCAGCGCGCGCG 129865
QY 548 CTACGCTCAGGTGCGCGCGCAGCATGTATCCACGCTGTCTCCCGCACCACCGCGCG 607
Db 129866 CTTCTGCTGCGCGCGCGCGCGCGCGCGT-----CCCGCGCAGGCGCATCAGCACCGCGC 129920
QY 608 CGGGCCAAAGGATCGCGCGCGTGTGAGGCTTACGAAGCGCGCGGAGTACGCGCGGGGTGCC 667
Db 129921 GTGCGCGACGAACCGCAGCTCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 129980
QY 668 GATGATGTCCCGCTGTGCGAGCCC 692
Db 129981 CGGGCG 130005

RESULT 9
AAD25519/c
ID AAD25519 standard; DNA; 154746 BP.

XX AAD25519;

XX 26-MAR-2002 (first entry)

DE Human herpesvirus 2 complete DNA genome. .

XX Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
KW antibacterial; fungicide; protozoicide; antirheumatic; antiinflammatory;
KW antiarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW vasculitis; ds.

XX Human herpesvirus 2.

PN WO200176643-A1.
PD 18-OCT-2001.
PP 06-APR-2001; 2001WO-US11372.
PX 07-APR-2000; 2000US-195680P.
PY (BAYU) BAYLOR COLLEGE MEDICINE.
PA Orson FM, Kinney BM, Bhogal BS;
PI WPI; 2002-066308/09.
PR
PT Composition for oral delivery of vaccines, comprises expression vector
PT containing antigenic genomic sequence, bound to aggregated
PT protein-polycationic polymer conjugate or suspension
PS Disclosure; Page 90-132; 145pp; English.
PX The invention relates to a composition comprising an expression vector
CC bound to an aggregated protein-polycationic polymer conjugate or
CC suspension. The expression vector contains a promoter polynucleotide
CC sequence operatively linked to a polynucleotide sequence encoding an
CC antigen which is a fragment of a gene or genome associated with an
CC infectious disease, cancer and autoimmune disease such as rheumatoid
CC arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
CC consisting of bacterium, fungus, protozoa and virus such as human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
CC virus (HCV), influenza and respiratory syncytial virus (RSV), and
CC optionally comprising a nucleotide sequence encoding a cytokine (or a
CC cytokine expression vector), is useful for inducing an immune response
CC (systemic and/or mucosal) in an organism. The cytokine expression vector
CC contains a sequence for granulocyte macrophage-colony stimulating factor
CC (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
CC the antigen and the cytokine are under transcriptional control of same or
CC different promoter polynucleotide sequences. The expression vector, as a
CC DNA vaccine is useful for treating a condition in an organism. The
CC present sequence is human herpesvirus 2 complete DNA genome related
CC to the invention.
XX
SQ Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 other;
Query Match 6.7%; Score 63.4; DB 24; Length 154746;
Best Local Similarity 45.8%; Pred. No. 0.027;
Matches 259; Conservative 0; Mismatches 301; Indels 5; Gaps 1;
QY 128 GCGCGCGCGCGCGCGCTGTGCTTCCGCGCCCTGTGCGCGCTGTGCGCGCTGTGCGCGCGA 187
DB 152296 GCGCGCGCGCGCGCGCTGTGCGCGCCCTCGCGCGCGCGCGCGCGCGCGCGCG 152237
QY 188 GGTGAGCGCGCGAGGCTCAGGACCTGAGTCTGCGAGGAGTCCAGAGAGCCCGCT 247
DB 152236 GCGCGCGCGCGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGG 152177
QY 248 CGACGCGTGTGCGCGAGTCTCTGACCGCGAGCTAACCGCGCGCGCGCGCGCGCGCT 307
DB 152176 CGGCGCGCGCGCGCGCGGTCTCGCGCGGCTCTTCTTGGGGGGCGCGCGCGCGCG 152117
QY 308 TGGCCCGTTCCGTTGGGGGACCGCGGCTCCGCGATGCGGTGTGCGAGCTCCAGGACGT 367
DB 152116 GCGCGCGCTTGGCGCGCGCGGCTCTTGGCGTTCGCGCTCCGCGCGCGAGCGG 152057
QY 368 GAGCGCGAGTGCCTGTGCGCGCCATCCGAGCATGTTAGGGGTACAGAGAGCCAT 427
DB 152056 GCGCGCGAGAGTGTGCGCGCGCGCGCGCGTGTGCGCGAGCGAGCGCGCGCTTGT 151997
QY 428 GCGCGCGTGGAGAGAGTGTGTGGCCATGCGGGCGCGAGCAGCAGCCCGCGCGAGG 487
DB 151996 TGTGAGAGAGAGTGTCCGCGCGCGCGCGCGCGCGAGTCTAGCAGCGCGGCTCCG 151937
QY 488 AGGAGGAGGAGGACAGGGGGGCTTACTACTACCCCTCAGCGCGGCCAGGAGAGGATACGG 547

DB 151936 GCAGCGCGCGCGCCCGCGGCGACGAGGCTCAGCGCGCGCGCGCGCGCGCGG 151877
QY 548 CTACCGTTCAGGTCGCCAGCGCAGATGTATCCACCGTGTCTCCCGGCGACACCGCGCG 607
DB 151876 CTTCCGTCGCGCGCGCGCGCGCGT-----CCCGCGCAGCGCATCAGCAGCGC 151822
QY 608 CGGCGCAAGGATCGCGCGCGTGTAGCGTTACGAAGCGCCCGGAGTACGCCGCGGGTTGCC 667
DB 151821 GTCCGCGCAGAACCGCAGCTCCGCGAGCGCAGCGCGCGCGCGCGCGCGCGT 151762
QY 668 GATGATGTCCCGCGTGTGCGAGCCC 692
DB 151761 CGGCG 151737
RESULT 10
AAL40781
ID AAL40781 standard; DNA; 88421 BP.
XX
AC AAL40781;
DT 03-OCT-2002 (first entry)
XX
DE 88421nt genomic DNA of ramoplanin producing Actinoplanes sp.
XX
KW Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;
KW biosynthesis gene cluster; bioengineering; peptide synthetase module;
KW adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;
KW chlorinate; lipdopeptide; gene; ds.
XX
OS Actinoplanes sp.
FH Key Location/Qualifiers
CDS 2077..3078 /tag= a
FT /product= "Protein of ORF 1"
FT 3118..4032 /tag= b
FT /product= "Protein of ORF 2"
FT 4038..5048 /tag= c
FT /product= "Protein of ORF 3"
FT complement (6665..5814) /tag= d
FT /product= "Protein of ORF 4"
FT complement (7703..6693) /tag= e
FT /product= "Protein of ORF 5"
FT complement (9464..8130) /tag= f
FT /product= "Protein of ORF 6"
FT 9691..10761 /tag= g
FT /product= "Protein of ORF 7"
FT complement (12751..10829) /tag= h
FT /product= "Protein of ORF 8"
FT complement (13617..12802) /tag= i
FT /product= "Protein of ORF 9"
FT complement (15203..13614) /tag= j
FT /product= "Protein of ORF 10"
FT 15591..15863 /tag= k
FT /product= "Protein of ORF 11"
FT 15880..19035 /tag= l
FT /product= "Protein of ORF 12"
FT 19032..39713 /tag= m
FT /product= "Protein of ORF 13"
FT 39713..65800

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2003, 14:01:53 ; Search time 75 Seconds
(without alignments)
5590.853 Million cell updates/sec

Title: US-10-053-410-3
Perfect score: 950
Sequence: 1 aaaaaaacccctcgtcat.....aaaaaaaaaaaaaaaaaaaaa 950

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91.2	9.6	2296	2	US-08-899-336-2
2	59.4	6.3	4524	2	US-08-845-998-7
3	59.4	6.3	4524	3	US-09-206-537-7
4	59.4	6.3	4524	3	US-09-430-854-7
C 5	59	6.2	3056	4	US-09-016-434-1153
C 6	57.4	6.0	3169	4	US-09-820-312D-713
C 7	57.4	6.0	4190	3	US-08-924-345-1
8	56.4	5.9	1556	3	US-09-043-937A-3
9	55.4	5.8	4403765	3	US-09-103-840A-2
10	55	5.8	4411529	3	US-09-103-840A-1
11	53.4	5.6	530	3	US-08-758-662-4
C 12	52.8	5.6	672	4	US-09-252-991A-14411
C 13	52.8	5.6	888	4	US-09-252-991A-14356
14	52.8	5.6	894	4	US-09-252-991A-6276
15	52.8	5.6	1530	4	US-09-252-991A-14293
16	52.8	5.6	3594	4	US-09-252-991A-14262
17	52.6	5.5	8438	1	US-07-945-283-1
18	52.4	5.5	1401	4	US-09-252-991A-14814
19	52.4	5.5	1857	4	US-09-252-991A-7125
20	52.4	5.5	1908	4	US-09-252-991A-15075
21	52.4	5.5	2925	4	US-09-252-991A-7069
22	52.4	5.5	3900	4	US-09-252-991A-14550
C 23	52.2	5.5	152331	3	US-09-128-155-16
C 24	52.2	5.5	4403765	3	US-09-103-840A-2
C 25	52.2	5.5	4411529	3	US-09-103-840A-1
C 26	51	5.4	525	4	US-09-252-991A-15328
C 27	51	5.4	714	4	US-09-252-991A-1104

28	5.4	759	4	US-09-252-991A-982	Sequence 982, App
C 29	5.4	957	4	US-09-252-991A-15376	Sequence 15376, A
30	5.4	963	4	US-09-252-991A-15461	Sequence 15461, A
C 31	5.4	1119	4	US-09-252-991A-3537	Sequence 3537, Ap
C 32	5.4	1289	2	US-08-344-833-1	Sequence 1, Appli
C 33	5.4	1467	4	US-09-252-991A-1142	Sequence 1142, Ap
34	5.4	1581	4	US-09-252-991A-3549	Sequence 3549, Ap
C 35	5.4	1947	4	US-09-252-991A-3557	Sequence 3557, Ap
36	50.8	966	4	US-09-252-991A-14511	Sequence 14511, A
C 37	50.8	969	4	US-09-252-991A-14770	Sequence 14770, A
38	5.3	1863	4	US-09-252-991A-3225	Sequence 3225, Ap
C 39	5.3	1926	4	US-09-249-585A-2	Sequence 2, Appli
40	5.3	1926	4	US-09-410-399-3	Sequence 3, Appli
C 41	5.3	2154	4	US-09-252-991A-3346	Sequence 3346, Ap
42	5.3	2580	3	US-09-050-863-2	Sequence 2, Appli
C 43	5.3	2580	4	US-09-359-081-2	Sequence 2, Appli
C 44	5.3	5452	2	US-09-130-114-1	Sequence 1, Appli
C 45	5.3	8705	4	US-09-647-344A-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1
US-08-899-336-2
; Sequence 2, Application US/08899336
; Patent No. 5955649
; GENERAL INFORMATION:
; APPLICANT: HIROTA, NAHIKO
; APPLICANT: KIHARA, MAKOTO
; APPLICANT: KURODA, HISAO
; APPLICANT: ITO, KAZUTOSHI
; TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA,
; TITLE OF INVENTION: GENE EXPRESSION CASSETTE, EXPRESSION VECTOR AND TRANSGENIC
; TITLE OF INVENTION: PLANT
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,336
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP HEI 8-193433
; FILING DATE: 23-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-0061-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2296 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-899-336-2

Query Match 9.6%; Score 91.2; DB 2; Length 2296;

```
Best Local Similarity 61.4%; Pred. No. 8.2e-09;
Matches 178; Conservative 0; Mismatches 88; Indels 24; Gaps 1;

QY 207 AGGACCTGAGTGTGCGAGGAGTCCAGAGAGCCCGCTCAGCGGTCCGCGCAGGTC 266
Db 145 AGGAGCTACAGTGTGAGCGGAGTCCAGAGAGTCCGTCAGAGGCGTCCGCGCGGTC 204
QY 267 CTCAGCCGCGAGCTAACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 326
Db 205 GTGACCAACAGCTG-----GTTGCGCCAGCTGCCATGGAGC 240
QY 327 ACCGGGTCGCGAGTGTGCGAGAGTCCAGAGAGTCCAGCGGAGTCCGCGTGC 386
Db 241 ACCGGGTCGAGATGAGTGTGCGAGAGTCCAGCGGAGTCCAGCGGAGTCCGCGG 300
QY 387 GCGCCATCCGAGAGTGTGCGAGGAGTCCAGAGAGTCCAGCGGAGTCCAGAGAGG 446
Db 301 GTGCGCTCAGCCAGGTCGTGAGGCAATACGAGCAGCAACCGAGGTGCCATCCAGGGA 360
QY 447 TGGTGGCCATGGGGCGGCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496
Db 361 GGAATCTTACCGCGGCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 410

RESULT 2
US-08-845-998-7
; Sequence 7, Application US/08845998
; Patent No. 5879892
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coulie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; CORRESPONDENCE ADDRESS: 16
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,998
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-2441
; TELEFAX: (617)720-2441
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4524 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 174..1433
; US-08-845-998-7

Query Match 6.3%; Score 59.4; DB 2; Length 4524;
Best Local Similarity 48.8%; Pred. No. 0.0072;
Matches 189; Conservative 0; Mismatches 196; Indels 2; Gaps 1;

QY 145 TGTGCTTCGCGCCCTGTGTGGCGGTGGCGGTCTGCGAGAGAGTCCAGAGAGTCCGCGCAGG 264
Db 532 TCTCTCTCGCGCTGCTCGCGCTCATGGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 591
QY 205 TCAGGAGACTGACAGTGTGCGAGAGTCCAGAGAGTCCAGAGAGTCCAGAGAGTCCGCGCAGG 264
Db 592 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 651
QY 265 TCTCTCACCGCAGACTAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 324
Db 652 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711
QY 325 GCACCGCGCGTCCGAGTGTGCGAGAGTCCAGAGAGTCCAGAGAGTCCAGAGAGTCCGCGT 384
Db 712 GCGCGCGCGTCTCTGCGCGCGTCCGCGCACCTCACCGCATATGCACAGCTTGGCGCAC 771
QY 385 GCGCGCGCGTCCGAGAGTGTGCGAGAGTCCAGAGAGTCCAGAGAGTCCAGAGAGTCCGCG 444
Db 772 TGTGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 829
QY 445 GCTGTGGCCATGGGGCGCGCGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 504
Db 830 GCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 889
QY 505 GGGGCTACTACTACCCCTGCAGCGCGC 531
Db 890 GCAGGTGGCAGCGGCATCGCGCGCGC 916

RESULT 3
US-09-206-537-7
; Sequence 7, Application US/09206537
; Patent No. 6130052
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coulie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/206,537
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,998
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
```



```

; ZIP: 94304
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1153:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3056 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g1524108
; US-09-016-434-1153

```

[illegible]

RESULT 6
US-09-620-312D-713/C
; Sequence 713, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui

```

; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungui
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 713
; LENGTH: 3169
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (116)..(2146)
US-09-620-312D-713

Query Match      6.0%; Score 57.4; DB 4; Length 3169;
Best Local Similarity 47.9%; Pred. No. 0.016;
Matches 195; Conservative 0; Mismatches 211; Indels 1; Gaps 1

Qy      122  CGCGCGCGCGCGCGCGCGCTGTGCTTCGCGGCCCTGTGTGCGCGCTGTGTGCGCGCTGTGCCTGCGCA 181
Db      551  CGCGCGCGTGGCGCGCCCTTGCCTGCGCTGTGCTGTGCGCGTGTGCTGTGCGGTGCTGCT 492

Qy      182  AGCGAGGTCGAGCGCGCAGAGGCTCAGGACCTCGAGTGTCTGTGCGAGGAGTTCAGAGAG 241
Db      491  GCGGCTCGCGCGCGCGCGCGCGCGCGCTGCTCGCGCGCGCGCGCGCTTCTTCGCGCA 432

Qy      242  CCGCTCGAGCGGTGCGCGCAGGCTCTCGAGCGGACGTAACCGCGCGCGCGCGCGCGG 301
Db      431  CGTCATCGCGCGCGCGCGCGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372

Qy      302  CGGCGTTGGCGCGTTCGCGTGGGCGACCGGCGCTCCGATGCGGTGCTGCCAGCAGCTCCA 361
Db      371  CGCCCTTGAGCCCTTTGTTTGSCCGAGGCGCATGCGCGCGCGCGCGCGCGCTCGGGCG 312

Qy      362  GAGCGTGAGCGCGGAGTGCCGCTGCGCGCGCGCATCGGAGCAGTGTGTAGGGGCTACGAGGA 421
Db      311  CGGCGCTGGGAGGCGAGCGAGCGCGCGCGCGCGCGCGCGGAACTTCAGGGGCTCCCCCG 252

Qy      422  GGCGATCCCGCGTGTGAGAAAGGCTGTGTGCCATGCGGGCGCGAGCAGCAGCGCGCGCC 481
Db      251  GACTGGCGGAGCGGCACCGCGGCGCTGCTGG-CGTCGAGCGCGAGGAGCGCGCGCC 193

Qy      482  GCAGGGAGGAGGAGGACAGGGGGGCTACTACTACCCCTGCAGCC 528
Db      192  GAGCGGTGGGACACCTTCTCCGAGGGGCTCGCGCGCGCGTGTGGCGCGCC 146

```

RESULT 7
US-08-924-345-1/c
; Sequence 1, Application US/08924345
; Patent No. 6224878
; GENERAL INFORMATION:
; APPLICANT: LEUNG-TACK Patricia
; APPLICANT: LEGASTELLOIS Isabelle, Christine, Marie-Andree
; APPLICANT: AUDONNET Jean-Christophe, Francis
; APPLICANT: RIVIERE Michel, Emile, Albert
; TITLE OF INVENTION: Mutants and vaccines of the Infectious
; TITLE OF INVENTION: Bovine Rhinotracheitis virus
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LARSON AND TAYLOR
;


```
Query Match          5.9%; Score 56.4; DB 3; Length 1556;
Best Local Similarity 48.4%; Pred. No. 0.024;
Matches 156; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 101 CTGAGAAGAGATGGCTAAGATCCCGCGCGCGCGCGCGCGCGCTGTGCTTCCGCGCCCT 160
DB 740 CGGCAAGGCGGTGGCGAGCAGCGCGCTGGAGGCGCGCGCGCGCGCGCTGGACCT 799

QY 161 GGTGGCGGTGGCGCTGTGCCAAGCAGAGCTGAGCGCGCAGAGCTCAGGACCTGCAATG 220
DB 800 GGTGGTGGTGAACCCCTGTGTGTGGGCCCCCTGCTGCAGCGCAGCGTGAACCCAG 859

QY 221 CTGGCAGGAGGTCCAGAGAGCCGCTCGACGCGTGCAGCGCTCTCGACCGCAGCT 280
DB 860 CATCGCGCACATCTCAATGACTGACGCTCGGCGCGCACCTTTCGCCAAGCCCTGCA 919

QY 281 AACCGCGCGCGCGCGCGCGCGCTGTGCGCGCTTCCGCTGGGCGACCGGGTCCGGAT 340
DB 920 GCGGTACGTGGACGTGCGCGACGTGCGCAGCGCACCTCCCGCTTTCGAGAGCCCCG 979

QY 341 GCGGTCTGCCACAGCTCAGAGAGCTGAGCGCGAGTGCCTGCGCGCCATCCGGAG 400
DB 980 CCGCTCCGCGCCACCTCTTCGCCCGAGCGCTCTCCACCGCGAGGAGCTCGTCCGCAT 1039

QY 401 CATGGTCAGGGGTACGAGGAG 422
DB 1040 CCTCGCAAGCTTCTCCCGAG 1061

RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          5.8%; Score 55.4; DB 3; Length 4403765;
Best Local Similarity 45.9%; Pred. No. 0.076;
Matches 230; Conservative 0; Mismatches 266; Indels 5; Gaps 1;

QY 69 CGACAGACAAAGATTTGTGAGGCTGATCCCGCTGAGAGAGATGGCTAAGATCGCGCG 128
DB 3941279 CCACAGGCAACGCGGTGTGCGGGGGCGCGCGCCAGCGGTGACGGCGAGCGCGCG 3941338

QY 129 GCGCGCGCGCGCGCTGTGCTTTCGCGCGCTGTGCGCGCTGTGCGCGCTGTGCGCGAG 188
DB 3941339 GGGCGCGCGCGCGACCGACCGCGCGCCACCGCGCGCACCGGTTCGCGCGTGGAG 3941398

QY 189 GTCAGCGGCGAGGCTCAGGACCTGAGTGTGCGAGAGGTCCAGAGAGCCCGCTC 248
DB 3941399 -----CCGCGCGGCGCGCGCGCGCGCGCGCGGTAAACAGCGGTGCGCGCGCAC 3941453

QY 249 GACGCGTGCAGGCTCTCTCGACCGGACGCTAACCGCGCGCGCGCGCGCGCGCGCTT 308
```

```
DB 3941454 CAACGGCTCCGCGCGCGCGCGCGCACTCGCGGACAAAGCGCGCGCGGTGCTGGCGG 3941513
QY 309 GGCCTGTTCCGCTGGGACACCGGGTCCGGATGCGGTGCTGCCAGACAGCTCCAGGACGTG 368
DB 3941514 GGCCTGCGCGGATACCCACCGGCATCGCGCGCACCGCGGTGACCGCGGCGCACCGGCGG 3941573
QY 369 AGCCGCGAGTGCCTGCGCGCCCATCGGGAGCATGGTTCAGGGGTCTACGAGGAGGCCATG 428
DB 3941574 AGCGCGGAGCGCGCGGGCGCGGAGCGGCCCGGACCGGAGGACCGCGGCGCATGAT 3941633
QY 429 CCGCGCTGTGAGAAAGTGTGGCCATGCGGGCGCGCAGCAGCAGCCGCGCGCGAGGA 488
DB 3941634 CGGCACCAACAGCAACGCGGTGTGCGGGGCGCGCGCGCAAGCGGTGACCGCGGAGC 3941693
QY 489 GGAGGAGGAGGACAGGGGGGGTCTACTACTACCTTCAGCGCGCGCAGGAGGATACGCG 548
DB 3941694 CCGCGGGCGCGCGCGCGCGACAGCCCGCGCGCACCGCGCGGACCGGGTTTCGCGG 3941753
QY 549 TACGCTCAGGGTGGCGCAGCGG 569
DB 3941754 TCGAGCGCGCGGGCGCGCGG 3941774

RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          5.8%; Score 55; DB 3; Length 4411529;
Best Local Similarity 47.9%; Pred. No. 0.09;
Matches 218; Conservative 0; Mismatches 235; Indels 2; Gaps 2;

QY 113 GGCTAAGATCGCGCGCGCGCGCGCGCTGTGCTTTCGCGCGCTGTGCTGGCGCTGGC 172
DB 3947296 GGCCTAAGCGGTGACGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3947354

QY 173 CGTCTGCCAAGCGCGAGTTCGAGCGCGCAGAGGCTCAGGAGCCTGCGAGTGTGCGCAGAGT 232
DB 3947355 CACCGCGCGCACCGGGTTTCG-CCGGTGGAGCGCGCGCGCGCGCGCGCGCGCGCG 3947413

QY 233 CCAGGAGAGCCCGCTCGACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 292
DB 3947414 GCGGTGCCGCGCGCACCAACGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3947473

QY 293 CCGCGCGCGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 352
DB 3947474 CCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3947533

QY 353 GCAGTCTCCAGGAGCGTGGAGCGCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 412
DB 3947534 ACAGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3947593

QY 413 CTACGAGGAGGCGCATGCGCGCGCGCTGAGAAAGGCTGGTGGCGCATGGGGCGCGCAGCA 472
DB 3947594 GCACCGCGCGCATGATCGCACCAAGGCAACCGCGGTGTCGCGCGCGCGCGCGCG 3947653
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2003, 15:12:39 ; Search time 353 Seconds
(without alignments)
8866.730 Million cell updates/sec

Title: US-10-053-410-3

Perfect score:

Sequence: 1 aaaaaaacccctcgat.....aaaaaaaaaaaaaa 950

Scoring table: IDENTITY NUC

Gapop 10.0 ; Gapext 1.0

Searched: 2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

100% Processing: Minimum Match 0%
Maximum Match 100%

Maximum match 100%
Listing first 45 summaries

Database : Published Applications NA:*

```

1: /cgn2_6/pdata/2/pubna/us07 PUBCOMB.seq.*
2: /cgn2_6/pdata/2/pubna/PCT_NEW PUB.seq.*
3: /cgn2_6/pdata/2/pubna/us06 NEW PUB.seq.*
4: /cgn2_6/pdata/2/pubna/us06 PUBCOMB.seq.*
5: /cgn2_6/pdata/2/pubna/us06 NEW PUB.seq.*
6: /cgn2_6/pdata/2/pubna/PCTUS PUBCOMB.seq.*
7: /cgn2_6/pdata/2/pubna/us08 NEW PUB.seq.*
8: /cgn2_6/pdata/2/pubna/us08 PUBCOMB.seq.*
9: /cgn2_6/pdata/2/pubna/us09A PUBCOMB.seq.*
10: /cgn2_6/pdata/2/pubna/us09B PUBCOMB.seq.*
11: /cgn2_6/pdata/2/pubna/us09C PUBCOMB.seq.*
12: /cgn2_6/pdata/2/pubna/us09 NEW PUB.seq.*
13: /cgn2_6/pdata/2/pubna/us10A PUBCOMB.seq.*
14: /cgn2_6/pdata/2/pubna/us10B PUBCOMB.seq.*
15: /cgn2_6/pdata/2/pubna/us10 NEW PUB.seq.*
16: /cgn2_6/pdata/2/pubna/us06 NEW PUB.seq.*
17: /cgn2_6/pdata/2/pubna/us06 PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
C 1	64.6	6.8	6.8	1614	10	US-09-376-740-45	Sequence 45, Appl
C 2	64.6	6.8	6.8	1614	13	US-10-023-529-45	Sequence 45, Appl
C 3	64.6	6.8	6.8	1614	13	US-10-023-523-45	Sequence 45, Appl
C 4	64.6	6.8	6.8	1614	10	US-09-376-740-50	Sequence 50, Appl
C 5	64.6	6.8	6.8	1425	13	US-10-023-529-50	Sequence 50, Appl
C 6	64.6	6.8	6.8	13425	13	US-10-023-523-50	Sequence 50, Appl
C 7	63.4	6.7	3957	12	US-10-200-562-193	Sequence 193, App	
C 8	63.4	6.7	3957	12	US-10-237-551-193	Sequence 193, App	
C 9	63.4	6.7	154746	12	US-09-827-688-8	Sequence 8, Appl	
C 10	63.4	6.7	154746	12	US-09-827-688-8	Sequence 8, Appl	
C 11	63	6.6	88421	10	US-09-376-059-1	Sequence 1, Appl	
C 12	61.2	6.4	3133	12	US-10-017-161-1483	Sequence 1483, Ap	
C 13	59.4	6.3	3824	12	US-10-116-275-197	Sequence 197, App	
C 14	59	6.2	3035	10	US-09-864-864-311	Sequence 311, App	
C 15	59	6.2	3393	12	US-09-814-353-19395	Sequence 1995, A	
C 16	57.4	6.0	3169	14	US-10-037-270-713	Sequence 713, App	

Query Match 6.8%; Score 64.6; DB 10; Length 1614;
Best Local Similarity 46.4%; Pred. No. 7.2e-06;

QY 108 GAGATGGCTAAGATCGCCGGCGGCGGCGGCGCTGTGCTTCGCGGCCCTGGTGGCC 167

ALIGNMENTS

RESULT 1

US-09-976-740-45/c
; Sequence 45, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03

Query Match 6.8%; Score 64.6; DB 10; Length 1614;
Best Local Similarity 46.4%; Pred. No. 7.2e-06;

QY 108 GAGATGGCTAAGATCGCCGGCGGCGGCGGCGCTGTGCTTCGCGGCCCTGGTGGCC 167

Qy	548	CTACGGTCAGGTCGGCCAGCGGCAAGATGATACACCGTGTGTCCTCCGGGCAACCAACCGGG	607
Db	2167	CCTCGTCTGCCCGGCGCCAGCGCGAGGT-----CCCCGGCAGGCGCATACACACACGCG	2113
Qy	608	CGGGCCAAAGATCGGGCGTGTGAGGCTTACGAAGCCCGGAGATACGCGCGGGGTTGCC	667
Db	2112	GTCGCGACGAACCGCAGCTGCGCGACACCAAGCGCGAGCGGGGCGCTGGCGTCGG	2053
Qy	668	GATGATGTCCGGCTGTCCGAGCCC	692
Db	2052	CGGGGCGGCGCGCGCCCGGCGCCC	2028

```

RESULT 8
US-10-237-551-193/c
; Sequence 193, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 3957
; TYPE: DNA
; ORGANISM: HSV2
US-10-237-551-193

```

Query Match	5.7%;	Score 53.4;	DB 12;	Length 3957;
Best Local Similarity	45.8%;	Pred. No. 1.8e-05;		
Matches 259;	Conservative 0;	Mismatches 301;	Indels 5;	Gaps 1;
Qy	128	GGCGGCGCGCGCGCTGTGTCTTCGCGGCCCTCGTGCGCGCTGTGCGCGCTCTGCCAAGCGGA	187	
Db	2587	GGCGGCCACAGCCGCTTGGGGTCTGGGGCCCTCGCGGGCGCGCGGTCAGCCGCCGG	2528	
Qy	188	GGTCGAGCGCGAGAGGCTCAGGGACCTGCAGTGTCTGCAGAGAGTCCAGAGAGCCCGCT	247	
Db	2527	GGCGCGCGCGCGCGCGCGCTCGCGGGCGGGGGCGCGGCCCGCGCGAGGG	2468	
Qy	248	CGAGCGTTCGCGCCAGGTCTCGACCGCGAGCTAACCGCGCGCGCGCGCGCGCGCGCGT	307	
Db	2467	CGGCGCGCGCGCGGGGGCGTCCGCGCGGCTCTTCTTCGGGGGGCGCGGGCGCGCGCG	2408	
Qy	308	TGGCGCCGTTCCGGTGGGGCACCGGGCTCCGGATCGGTCGTCTGCAGCAGCTCCAGGACGT	367	
Db	2407	CGGCGCCCTGGCGCGGGGGGGCTCTTGGCTTGGCGGCTCTCCGCGCGCGGAGCGGG	2348	
Qy	368	GAGCGCGAGTGCCTGCGCGCCATCCGAGAGATGGTCAGGGGCTACGAGGAGGCCAT	427	
Db	2347	GGCGCGGAGCGAGTTCGCGCGCGCGCAGCGGTCTCGGCACGAGGGGCGCAGGCTCTGGT	2288	
Qy	428	GCCECCCTGGAGAAAGCTGTGGCCATGGGGGCGCGAGCAGACGCGCGCGCGCAGGG	487	
Db	2287	TCTGGAGAGCAGGTCGCGCGCGCGCGCGCGGAGCTCAGCAGCGCGGGCTCGCGG	2228	
Qy	488	AGGAGGAGGAGGACAGGGGGCTACTACTACCCCTGCAGCGCGCCAGAGAGGATACGG	547	
Db	2227	GCAGCGCGGGCCACAGGGCCCCGGCGACAGGGTCA CGCGCGCACCGCGGCACGGCGG	2168	
Qy	548	CTACGGTCAAGGTCGCGCAGCGCAGATGTATCCACCGTGTCTGCCCGCACCCCGCGG	607	
Db	2167	CCTCGCTCCGCGCGCCACGGCAGGT-----CCCCGGCAGGGCGCATCAGCACCGGC	2113	
Qy	608	CGGGCCAAAGANTCGGCGCGGTGAGGGCTTTACGAAGCCCGGGAGTACGCCCGCGGGTTGCC	667	
Db	2112	GTGCGCACGAACCGCAGCTCGCGCAGCGCGCGAGGGCGGGCGCTCGCGCTGCGG	2053	

Qy 668 GATGATGTCCGGCTGTCTGGAGCCC 692
Db 2052 CGGGGCGGGCCCGCCCGCGGGGCCCC 2028

RESULT 9
US-09-827-688-8
; Sequence 8, Application US/09827688
; Publication No. US20030165476A1
; GENERAL INFORMATION:
; APPLICANT: ORSON, FRANK
; APPLICANT: KINSEY, BERNA
; APPLICANT: BHOGAL, BALBIR
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION
; TITLE OF INVENTION: AGENTS
; FILE REFERENCE: P01949US1/10004014
; CURRENT APPLICATION NUMBER: US/09/827,688
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,680
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Query Match	6.7%	Score 63.4	DB 12	Length 154746
Best Local Similarity	45.8%	Pred. No. 5.4e-05		
Matches 259	Conservative 0	Mismatches 301	Indels 5	Gaps 1
Qy	128	GGCGGCGGCGCGCGCTGTGCTTCGCGGCCCTGTGTCGCCGTGCGCGTGTCCAAGGCGA	187	
Db	129446	GGCGGCGCGAGCCGCTTGCGGGTCGGGGCCCTCGCGGGCGCGCGGTTCAGCGCCGCGG	129505	
Qy	188	GGTTCGAGCGCGAGAGGCTCAAGGACCTTCAGTGTGTGCGAGAGGTTCAGGAGAGCCCGCT	247	
Db	129506	GGCGGCGCGCGCGCGCGCGCTCGCGGGGCGGGGGCGCGCCCCCGCGGAGGGG	129565	
Qy	248	CGAGCGGTGCGCCAGGTCTTCGACCGCAGCTAACCGCGCGCGCGCGCGCGCGGT	307	
Db	129566	CGGCGCGGGCGGGGGCGGTTCGCGCGGCTCTTCTTCGGGGGGCGCGGGCGCGCCCG	129625	
Qy	308	TGGCCCCCTTCGCGTGGGGCGACCGGGCTCCGGATCGGTGCTGCGAGAGCTCCAGAGAGCT	367	
Db	129626	CGGCGCCCTTGGCGGGGCGGGGCTCTTGCGCTTGCGCGCTCCGCGCGCGGGAGCGG	129685	
Qy	368	GAGCGCGAGTCCGCTGCGGCGGCCATCCGGAGGATGGTCAGGGGCTACGAGGAGGCCAT	427	
Db	129686	CGCGGGCGAGCGAGTTCGCGCCCGCGCGAGCGGTGTCGGCGAGCAGGGGCGCAGGCTCTGGT	129745	
Qy	428	GCGCCCGCTCGAGAAAGGCTTGTGGCCATGGGGCGCGCAGCAGACGCGCGCGCGCAGG	487	
Db	129746	TCTGGAAGACAGTTCGCGCGCGCGGGCGGGCGGAGCTCAGCAGGCGCGGGTCCCGG	129805	
Qy	488	AGGAGGAGGAGGACAGGGGGGCTACTACTCCCTGACGCGCGCAGAGAGGAGATACGG	547	
Db	129806	GCAGCGCGGGGCCAGGGGCCCGCGCAGCAGGCTCAGCGCGCAGCAGCGCGCCACGGGG	129865	
Qy	548	CTACGGTTCAGGTGGCGCAGCGGAGATGTATCCACCGTGTCTCCCGCACCCCGGGG	607	
Db	129866	CCTCGCTTGGCCCGGCCACGGCGAGGT-----CCCCGCGAGGCGCATCAGCACCAGCGC	129920	
Qy	608	CGGGCCAAAGGATCGGCGCGGTGAGGCTTACGAAGCGCCGGGAGTACGCGCGGGGTTCGC	667	
Db	129921	GTGCGCGCAGAACCGCAGCTCGCGCAGCCAGCGCGCAGGCGCGGCGCGGTCTGGCGTGGG	129980	
Qy	668	GATGATGTCGGGCTGTTCGAGCCCC	692	
Db	129981	CGGGGCGGCGCGCGCGCGCGCCCC	130005	

OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (210)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (212)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (214)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (219)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (222)..(225)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (229)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (231)..(232)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (238)..(239)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (242)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (245)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (248)..(260)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (265)..(270)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (272)..(276)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (278)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (282)..(283)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (287)..(290)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (302)..(306)	
OTHER INFORMATION:	a, t, c, g, unknown or other
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (311)..(313)	
OTHER INFORMATION:	a, t, c, g, unknown or other

FEATURE:	
NAME/KEY: modified_base	
LOCATION: (323)..(334)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (336)..(341)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (345)..(363)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (365)..(368)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (370)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (374)..(375)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (379)..(385)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (387)..(388)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (390)..(393)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (395)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (397)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (400)..(402)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (409)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	

[illegible]

Db 574 GNNNGGGGGGGGGGNNNNNGGGGGGNNNGGGGGGNNNGGNNNGGGGGGGGG 633
QY 369 AGCCGCGAGTCCGCTCGCGCCCATCCGAGCATGTCAGGGCTACAGAGGAGGCATG 428
Db 634 GNNNGGGGGGGGNNNNNGGNGGGGGGGGGGNNNGGGGGGGGGGGGGGGGGGG 693
QY 429 CCGCCGCTGAGAAAGCTGTGTCATGGGCGGCGAGCAGCGCGCGCGCAGGGA 488
Db 694 GCGNNNGGGGNGGNGGGGGGGGGGNNNNNGGNGGGGGGGGGGNGGNGGGGGGG 753
QY 489 GAGAGGAGGACAGGGGGCTACTACTACCCCTGACCGCGCAGGAGGATACGC 548
Db 754 GGGGGGNNNGGGGGGGGNNNNNGGGGNGGNGGGGGGGGGGGGGGGGGNNNN 813
QY 549 TACGCTCAGGGTGGCCAGCGCAGATGTATCCACCGTGTCTCCCGCACACCGCGGC 608
Db 814 GGGGGGGGGGGGNNNNNGGGGGGGGNNNGGGGGGGGGGGGGGGGGGGGG 873
QY 609 GGCCCAAGATCGCCCGCTGAGGCTTACGAGGCCCGGAGTACCGCGGGGTTGCCG 668
Db 874 GCGGGGNGGNGGGGGGGGGGNNNNNGGGGGGGGGGGGNGGNGGGGGGGGG 933
QY 669 ATGATGTCCGCTGTCTCGAGCCCCAGGAGTG 700
Db 934 GGGGGGGGNNNGGNNNGGGGGGGGGGGGGGGGGGG 965

RESULT 13
US-10-116-275-197
; Sequence 197, Application US/10116275
; Publication No. US20030211476A1
; GENERAL INFORMATION:
; APPLICANT: Elan Pharmaceutical Technology
; APPLICANT: O'Mahony, Daniel J.
; APPLICANT: Brayden, David
; APPLICANT: Byrne, Daragh
; APPLICANT: Lambkin, Imelda
; APPLICANT: Higgins, Lisa
; TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and Methods and
; FILE OF INVENTION: Compositions Targeting Peyer's Patches and M Cell Receptors
; FILE REFERENCE: E1067/20087
; CURRENT APPLICATION NUMBER: US/10/116,275
; CURRENT FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 349
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 197
; LENGTH: 3824
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-116-275-197

Query Match 6.3%; Score 59.4; DB 12; Length 3824;
Best Local Similarity 48.8%; Pred. No. 0.00017;
Matches 189; Conservative 0; Mismatches 196; Indels 2; Gaps 1;
QY 145 TGTGCTTCGGGGCCCTGTGTCGCGTGTGCGCTGTGCGAAGCGAGGTGCGAGCGCAGAGGC 204
Db 593 TCTCTTCGCGCTGCTCGCGCTCATGGCCGCGGGGCGCGCGGGCGCGCGCGCGCG 652
QY 205 TCAGGACCTGCAAGTGTGTCAGAGAGTCCAGAGAGCCCGCTCGACGCGTGCCTCCAGG 284
Db 653 GCGGCGCGCCACGACGCGCCCGGGGGGGTTCGCGCGCGCGCGCGCGCGCGCGCG 712
QY 265 TCCTGACCGCAGCTAACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 324
Db 713 GCGGCGCGCGCGCGGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 772
QY 325 GCACCGGGCTCCGATCGCGTGTGTCAGAGCTCCAGAGCGTCCAGAGCGTCCGCT 384
Db 773 GCGGCGGGCTCTTGGGCGGCTCCGCGCACCTCACCGCATATGCACAGCTTGGGCGACC 832
QY 385 GCGCCGCCCATCCGAGCATGCTCAGGGGCTTACGAGGAGGCCATCCGCGCGCTCGAGAAAG 444

Db 833 TGTGCGACCCCGCGCGCGCGCGCGCATGAAATGCGCTCCGCGGTGCGCGC--ACCCCGG 890
QY 445 GCTGTGGGCATGCGGGCGCGCAGCAGCGCGCGCGCAGGAGGAGGAGGAGGAGGAGG 504
Db 891 GCTGTGGCGCGCGCGCGCGCGCACCAACGCGCGGCGGCGGCGGCGCGCGCGCTGCGCGCG 950
QY 505 GGGGCTACTACTACCCCTGACCGCGGC 531
Db 951 GCAGGTGGCAGCGGCATCGCGCGCGGC 977

RESULT 14
US-09-864-864-311/c
; Sequence 311, Application US/09864864
; Patent No. US20020102679A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Secrist, Heather
; APPLICANT: Lodes, Michael J.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steve P.
; APPLICANT: Mannion, Jane
; APPLICANT: Benson, Darin R.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.523
; CURRENT APPLICATION NUMBER: US/09/864,864
; CURRENT FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 311
; LENGTH: 3035
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-864-311

Query Match 6.2%; Score 59; DB 10; Length 3035;
Best Local Similarity 52.2%; Pred. No. 0.0002;
Matches 131; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
QY 290 CGCGCGCGCGCGCGGTGGCCGCTTCGCTGCGGCGCGCGCGCTCGGATGCGGTGCTG 349
Db 449 CGCGCGCAGCGAGCCCTCGCGCGCGCGCGCTGCGAGGCGCGCGCGCTGCTGCTG 390
QY 350 CCAGCAGCTCCAGGAGTGTGAGCGCGCGCTGCGCGCGCATCCGAGCATGCTGCTAG 409
Db 389 CCGAGGCTCGCGCGCGGAGCGCGCATTCACACGCGCGCGGGGGCGCCAGGGGGGAG 330
QY 410 GGGCTACAGGAGGCGCATCCCGCTGGAGAAAGCTGTGCGCCATGCGGGCGCGCAGCA 469
Db 329 GAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCAACTGCTGCGCGCT 270
QY 470 GCAGCGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 529
Db 269 GCG 210
QY 530 GCCAGGAGAGG 540
Db 209 CGCGCGGAGC 199

RESULT 15
US-09-814-353-19995/c
; Sequence 1995, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
 ;
 ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
 ; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER

; FILE REFERENCE: MRI-006B

; CURRENT APPLICATION NUMBER: US/09/814,353

; CURRENT FILING DATE: 2001-03-21

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031

PRIOR FILING DATE: 2000-03-21

PRIORITY FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/207,124

; PRIOR FILING DATE: 2000-05-25

; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: US 60/211,940; PRIOR FILING DATE: 2000-06-15
; PUBLICATION NUMBER: 00 00/211,340; PRIORITY DATE: 2000-08-13
; PRIORITY APPLICATION NUMBER: US 60/216,820

; PRIOR FILING DATE: 2000-07-07

; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661

;; PRIORITY FILING DATE: 2000-07-25

; PRIORITY FILING DATE: 2000-07-23
; PRIOR APPLICATION NUMBER: US 60/257,672

PRIOR FILING DATE: 2000-12-21

; ERROR LISTING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037

```

; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0

```

; SEQ ID NO 19995

```

; SEQ ID NO 13333
; LENGTH: 3393

```

LENGTH: 3393
TYPE: DNA

```

; TYPE: DNA
; ORGANISM: Homo sapiens

```

Query Match	6.2%	Score 59	DB 12	Length 3393
Best Local Similarity	52.2%	Pred. NO.	0.00021	
Matches 131	Conservative	0	Mismatches 120	Indels 0
				Gaps 0

290 CGCGCGCGCGCGCGTTCGGTGGGGCACCGGGCTCCGGATCGCGTCTG 349

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

[illegible]

QV 350 CCAGCAGCTCCAGGACGTGAGCCGCGAGTCCGCTGCCGCCATCCGGAGCATGGTCAG 409

bp 329 CCGAGGCTGCCGCCCGCGGAGGCCGCCATTCAACGCCGCCGGGGGGCCCGGGGGGGAG 270

DN 3Z9 CCGAGGCTGGCCACGAGGCCGCATTCACACGCGCAGCGCGCGAG 270

QY 410 GGGCTACGAGGAGGCCAATGCCCGCTGGAGAAAGGCTGGTGGCCATGGGGCGGCAGCA 469

269 GAGGGGCGGGGCCCGCGGCAGCGCGCTCGCGCCAACTGTGGCGCCGT 210

Ov 470 GCAGCCGCCGCGCAGGGAGGAGGACAGGGGGCTACTACTA CCCCCTGCAGCG 529

QY 470 GCAGCCCGCCG CAGGGAGGAGGAGGACAGGGGGCTACTACTACCCCTGCAGCCG 529

Search completed: November 29, 2003, 17:04:01
Job time : 371 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2003, 15:53:04 ; Search time 58 Seconds
(without alignments)
563.753 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAALCFALVAV.....MMCRLEPQECIFSGGDY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	156	13.6	434	23	AAW50386
2	130.5	11.4	157	19	AAW53261
3	117	10.2	362	22	AB362882
4	114.5	10.0	143	13	AB328122
5	110	9.6	385	22	AB371788
6	110	9.6	797	22	AB364072
7	109.5	9.5	426	14	AA331036
8	109	9.5	166	24	ABJ26643
9	106.5	9.3	174	22	AAV72901

10	106.5	9.3	1145	22	AAE09770	Drosophila melanog
11	105	9.1	694	23	AAU74823	Human REPTR 6 prot
12	105	9.1	694	24	AAE34057	FZD 8 protein. Un
13	105	9.1	694	24	ABU55903	Human protein Friz
14	103.5	9.0	248	22	AB559015	Drosophila melanog
15	103	9.0	202	22	AB562974	Drosophila melanog
16	103	9.0	655	23	AB571126	Mouse ischaemic co
17	101	8.8	217	20	AAW97318	A HOQO59 polypept
18	101	8.8	217	20	AAW97318	Human polypeptide
19	101	8.8	228	22	AAW41673	Human polypeptide
20	101	8.8	515	21	AA335806	Protein involved i
21	101	8.8	894	22	AB35148	Human NFAR-2 SEQ I
22	100.5	8.8	360	23	ABP63025	Human polypeptide
23	100.5	8.8	360	24	ABP97405	Human embryonic st
24	100	8.7	656	14	AA445555	Human EMS protein
25	100	8.7	656	19	AAW33813	Human EMS protein
26	100	8.7	665	22	ABG06460	Novel human diago
27	100	8.7	926	22	AB65135	Drosophila melanog
28	99.5	8.7	134	22	AAO05578	Human polypeptide
29	98.5	8.6	124	15	AA58636	Amylase inhibitor
30	98.5	8.6	124	18	AAW29525	Wheat amylase inhi
31	98.5	8.6	124	18	AAW24493	Novel amylase inhi
32	98	8.5	169	21	AB28185	Human heterogeneou
33	97.5	8.5	302	23	ABP63092	FLO11 gene expres
34	97.5	8.5	336	22	ABG12614	Novel human diago
35	97.5	8.5	474	21	AA27305	Arabidopsis thalia
36	97.5	8.5	477	21	AA27304	Arabidopsis thalia
37	97.5	8.5	533	21	AA27303	Arabidopsis thalia
38	97.5	8.5	671	22	ABG20586	Novel human diago
39	97	8.4	144	21	AA584357	Amino acid sequenc
40	97	8.4	148	15	AA47481	Polypeptide encode
41	97	8.4	297	24	ABR41261	Human DITP4 intrac
42	96.5	8.4	207	22	AB82383	Peanut allergen Ar
43	96	8.4	140	10	AA91891	Part of the sequen
44	96	8.4	173	21	AA584354	Amino acid sequenc
45	96	8.4	454	22	AB61917	Drosophila melanog

ALIGNMENTS

RESULT 1

AAW50386
ID AAW50386 standard; Protein; 434 AA.

AC AAW50386;

XX

DT 18-FEB-2002 (first entry)

XX

DE Wheat glutenin variant 1Ax2asteriskB.

XX

DE Wheat; glutenin; storage protein; variant; transgenic plant;

KW gluten.

XX

XX Triticum aestivum.

OS

PN WO200179477-A2.

XX

PD 25-OCT-2001.

XX

PF 13-APR-2001; 2001WO-HU00045.

XX

PR 14-APR-2000; 2000HU-0001563.

XX

PA (MAGY) MTA MEZOGAZDASAGI KI.

XX

XX Juhasz A., Tamas L., Bedo Z., Vida G., Karsai I., Lang L;

PI Tamasne Nyitrai E;

XX

XX WPI; 2002-017612/02.

DR N-PSDB; AA170852.

XX

PT A new nucleic acid comprising a variant allele of the 1Ax2asteriskHMW

PS Disclosure; SEQ ID NO 15438; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB101840-ABL16175), expressed DNA

CC sequences (AB101840-ABL16175) and the encoded proteins

CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 362 AA;

Query Match 10.2%; Score 117; DB 22; Length 362;

Best Local Similarity 39.8%; Pred. No. 0.0094;

Matches 33; Conservative 3; Mismatches 17; Indels 30; Gaps 7;

QY 103 EEAMPPLEKGM---WPW---GRQQP-----PQGGGGGGGGGGYPCSRPG-EGY 145

DB 107 QQPTTP---GWPPGQWAGAGNQNDGPDVQVQTGFAPGNWGGGGQ-----PQWNGP 155

QY 146 GYGQGGQRMYPPCRPTTGGGP 168

DB 156 GWNCGGGRG---PPRPGFNGGGP 176

RESULT 4

AAR28122

ID AAR28122 standard; Protein; 143 AA.

XX

AC AAR28122;

XX

DT 25-MAR-2003 (updated)

DT 17-MAR-1993 (first entry)

DE Alpha-amylase/trypsin inhibitor with OmpA signal peptide.

XX

KW Protein disulphide isomerase; RBI; disulphide bridge protein.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Peptide 1..21 /note= "OmpA signal peptide"

FT Peptide 22..143

FT /note= "RBI mature peptide"

XX

PN EP510658-A2.

XX

PD 28-OCT-1992.

XX

PF 23-APR-1992; 92EP-0106978.

XX

XX 26-APR-1991; 91DE-4113750.

XX

PA (BOE) BOEHRINGER MANNHEIM GMBH.

XX

PI Gluckshuber R, Wunderlich M, Skerra A, Rudolph R;

XX

XX WPI: 1992-358883/44.

DR N-PSDB; AAQ29910.

XX

XX Improving yield of natural conformation of di-sulphide bridge

PT proteins - obcd. by transformed prokaryotic host cells, by adding

PT thiol reagent, esp. glutathione, to culture medium

XX

PS Example; Page 10; 18pp; German.

XX

CC The sequence is that of alpha-amylase/trypsin inhibitor (RBI)

CC from *Eleusine coracana* with an OmpA signal peptide.

CC See also AAR28123 and AAR28124.

XX (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 143 AA;

Query Match 10.0%; Score 114.5; DB 13; Length 143;

Best Local Similarity 26.4%; Pred. No. 0.0058;

Matches 46; Conservative 17; Mismatches 64; Indels 47; Gaps 6;

QY 1 MAKIAAAALCPAALVAVAVCOGEVERQRLDLQCWQEVQESPLDACRQVLDRLTGG 60

DB 1 MKKTAIAIAVALAGFATVA---QASVTGTCIPGM---AIPHNPLDSCRWYVSTRTC-- 50

QY 61 GGGGGVGFPRMGTGLMRCCQQLQDVSRCCAAIRSMVRGYEEAMPPLKGMWPMGRQQ 120

DB 51 ---GVGPRLATQEMKARCCQLEAIPAYCRCEAVILMDGVVTSSGQHE-----GRLL 100

QY 121 QPPPGGGGGGGGGYPCSRPGEGYGGGQGRQMYP-----PCRPGTTGGGP 168

DB 101 QDLP-----GCPQVQRAFPAPKLIVTEVCNLAITHGGP 133

RESULT 5

ABB71788

ID ABB71788 standard; Protein; 385 AA.

XX

AC ABB71788;

XX

DT 26-MAR-2002 (first entry)

XX

DE *Drosophila melanogaster* polypeptide SEQ ID NO 42156.

XX

KW *Drosophila*; developmental biology; cell signalling; insecticide;

XX

OS *Drosophila melanogaster*.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI: 2001-656860/75.

DR N-PSDB; ABU15891.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from *Drosophila* and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Disclosure; SEQ ID NO 42156; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB101840-ABL16175), expressed DNA

CC sequences (AB101840-ABL16175) and the encoded proteins

CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 385 AA;

Query Match 9.6%; Score 110; DB 22; Length 385;
Best Local Similarity 29.7%; Pred. No. 0.046;
Matches 41; Conservative 7; Mismatches 50; Indels 40; Gaps 5;

QY 35 LQWQVQESPLDA---CRQLDRLQLTGGGGGGVPPFRWGTGLMRCCQLQDVSREC 90
DB 174 LQTHSIKNTLDVKAIKQDMDRQ---GGGGGGPRAGGRGQ-----216

QY 91 RCAAIRSMVRGYEAMPPLKKGWHPGRQOQPPQGGGGGGGYYPYPCSRPBGEGYGGG 150
DB 217 ---GDRGQGGG-----GNGGNRQNGGNWGGAGGGGGFGNGNFGGGGGGGSG 263

QY 151 GQRQMYPPCRPTGTTGGP 168
DB 264 GWNQ-----GGSGGP 275

RESULT 6
ABB64072
ID ABB64072 standard; Protein; 797 AA.
XX
AC ABB64072;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 19008.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WC200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL08175.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
XX
PS Disclosure; SEQ ID NO 19008; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-ABBS72072).
CC
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 797 AA;

Query Match 9.6%; Score 110; DB 22; Length 797;
Best Local Similarity 30.1%; Pred. No. 0.1;
Matches 41; Conservative 14; Mismatches 57; Indels 24; Gaps 7;

QY 38 WQEVQES-----PLDACRQVL-DRQLTGGGGGGV-GPPFRWGTGLMRCCQLQDVSREC 90

Db 658 WAERYRSVGKIEAEIAENTLANKQNGSGGSTNPSPQGGSGQ-----QQPNAAAAA 712
QY 91 RCAAIRSMVRGYEAMPPLKKGWHPW-----GROQPPPGGGGGGGG---YYPCSR 140
DB 713 AAAAAAGGAGGQSMPTQYAYQYQYAAAAAGGQPGQAFQPGGGGGGPPGNY 772
QY 141 PGEYGV--YGQGGGQ 154
DB 773 PGAGYGGYPAFGQQQ 788

RESULT 7
AAR31036
ID AAR31036 standard; Protein; 426 AA.
XX
AC AAR31036;
XX
DT 25-MAR-2003 (updated)
DT 26-MAY-1993 (first entry)
XX
DE C. acidovorans stereospecific hydrolase.
XX
KW S-(+)-2,2-dimethylcyclopropane carboxamide; 2,2-DMCPCA; prodn.;
KW cilastatin; antibiotic therapy; penem; carbapenem.
XX
OS Comamonas acidovorans A.18.
XX
FH Key Location/Qualifiers
FT Peptide 1..7
FT Peptide /note= "N-terminal peptide used to obtain AAO35116"
FT Peptide 13..18
FT Peptide /note= "peptide used to obtain AAO35117"
XX
PN EP524604-A2.
XX
PD 27-JAN-1993.
XX
PF 21-JUL-1992; 92EP-0112446.
XX
PR 26-JUL-1991; 91CH-0002247.
XX
PA (LONZ) LONZA AG.
XX
PI Zimmermann T, Robins K, Birch OM, Boehlen E;
XX
DR WPI; 1993-028686/04.
DR N-PSDB; AAO38949.
XX
XX Pure S-(+)-2,2-dimethylcyclopropane carboxamide prodn. by microorganisms(s) - with a stereospecific hydrolase isolated from PT microorganism, useful for inhibition of renal dehydropeptidase PT via cilastatin
XX
PS Disclosure; Fig 3; 19pp; German.
XX
CC The sequence is that of Comamonas acidovorans A.18 stereospecific hydrolase which can be used in the prodn. of optically pure S-(+)-2,2-dimethylcyclopropane carboxamide (2,2-DMCPCA), the starting material for the prodn. of cilastatin. Cilastatin inhibits e.g. renal CC dehydropeptidase (RDP) and in therapy it is administered with the CC antibiotic penem or carbapenem to prevent their inactivation by RDP.
CC (Updated on 25-MAR-2003 to correct PI field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 426 AA;

Query Match 9.5%; Score 109.5; DB 14; Length 426;
Best Local Similarity 22.8%; Pred. No. 0.057;
Matches 64; Conservative 16; Mismatches 70; Indels 131; Gaps 12;

QY 24 QGEVERQRRLDLCQWQEVQESPLD--ACRQLDRLQLTGGGGGGV-----66

Db 123 QAAADRRRLRRPPSRDHCPRQPLERFAMARGLVORLGRHGCGGAVLRAGHGRHLPLS 182
QY 67 -----GPFRTGTGLRNRCCQLODVSRCCRAAIRSM 98
Db 183 IGRQHHGAQHLGCGEPFRLTGRVGPGR-----PDGAQCCRC---RSH 226
QY 99 VRGYEAMP-----PLEK-----GWMP-----WGRQQQPPPGGGGGGG 133
Db 227 ARGHRRGGPAGPYGQPVORARLSCHDDARLLRPAGHPGPMGTGRRCCLPFGGAGPGG 286
QY 134 YYPCSRPGEGYGGGQGRQWYP--PCRPGTTGGG----- 167
Db 287 -----GAAPGQGRAGPLSRCHPG--GGGLAGAVRGDRRGARRHVPCTARGL 332
QY 168 -PRIGRVLRTKAREYAAGLPMMCR-----LSEPOECISFGG 203
Db 333 WPARVRDRPGAGPVHRHLPAAAAAPRGLHGFAGCTLRAGG 373
RESULT 8
ABJ26643
ID ABJ26643 standard; Protein; 166 AA.
XX
AC ABJ26643;
XX
XX 17-APR-2003 (first entry)
DT DT
DE DE
DE Rice seed allergen protein - SEQ ID No 48.
XX
XX Rice; Chrohn's disease antibody-binding peptide; Crohn's disease;
KW human vacuolar H+ transport ATPase; rice allergen;
KW human kruppel-like zinc finger protein 300.
XX
OS Oryza sativa.
XX
XX WO200288175-A1.
PN
XX
XX 07-NOV-2002.
PD
XX
XX 24-APR-2002; 2002WO-JP04061.
PF
XX
XX 24-APR-2001; 2001JP-0126121.
PR
XX 25-FEB-2002; 2002JP-0047384.
PR
XX (SAKA) OTSUKA PHARM CO LTD.
PA
XX
XX Saito H, Katsuragi K, Tachikawa T, Tanaka M, Ogino K, Taki T;
PI
XX
XX WPI; 2003-140169/13.
DR
XX
XX Crohn's disease antibody-binding peptide applicable in examination
PT reagents for use in diagnosis of Crohn's disease
PT
XX
XX Claim 8; Page 111; 120pp; Japanese.
PS
XX
XX The invention comprises Chrohn's disease antibody-binding peptides. The
CC peptides of the invention are useful for examining and diagnosing Crohn's
CC disease by recognising a human vacuolar H+ transport ATPase, or human
CC kruppel-like zinc finger protein 300, or rice allergen in a biological
CC sample from a subject. The present amino acid sequence represents a rice
CC polypeptide used in the invention.
XX
XX
SQ Sequence 166 AA;
Query Match 9.5%; Score 109; DB 24; Length 166;
Best Local Similarity 26.0%; Pred. No. 0.022;
Matches 39; Conservative 11; Mismatches 44; Indels 56; Gaps 6;
QY 46 LDACROVLDRQLTGGGGGGVGRFRTGLRMTCCQLODVSR--CRCAAIRSMVRGYEE 104
Db 52 LPRCAVVRKRCQVGHGAPGAVDEQ-----LRQDCRQLAAVDSDSWCRCSALNHV----- 102
QY 105 AMPLEKGNWPMCRQPPPGGGGGGGGGYFCSRPGEGYGGGQGRQWYPPCRPGTT 164

Db 103 -----GGIYRELGATDVGH-----PMAEVFGCRGD- 129
QY 165 GGGPRIGRVLRTKAREYAAGLPMMCELSEP 194
Db 130 -----LERA-----AASLPFACNVNDIP 146
RESULT 9
AAV72901
ID AAV72901 standard; Protein; 174 AA.
XX
XX AAV72901;
XX
XX 31-MAY-2001 (first entry)
DT DT
DE DE
DE Flax 2S storage protein.
XX
XX Flax; seed-specific promoter; storage protein; seed oil; vaccine;
KW protein expression; anticoagulant; cytokine; growth factor; pectinase;
KW interleukin; alpha-1-antitrypsin; anti-obesity protein; haemoglobin;
KW serum albumin; insulin; lactoferrin; myoglobin; pulmonary surfactant;
KW alpha amylase.
XX
OS Linum usitatissimum.
XX
XX
FH Key Location/Qualifiers
FT Region 126..136
FT /note= "Glutamine-rich region"
XX
XX WO200116340-A1.
PN
XX
XX 08-MAR-2001.
PD
XX
XX 25-AUG-2000; 2000WO-CA00988.
PF
XX
XX 27-AUG-1999; 99US-0151044.
PR
XX 27-OCT-1999; 99US-0161722.
PR
XX 30-MAY-2000; 2000CA-2310304.
PR
XX (SEMB-) SEMBIOSYS GENETICS INC.
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
XX Chaudhary S, Van Rooijen G, Moloney MM, Singh S;
PI
XX
XX WPI; 2001-226693/23.
DR N-PSDB; AAD03029.
DR
XX
XX Expressing non-native genes in flax seeds and seeds of other plant
PT species for altering the seed oil and protein composition in the seeds,
PT comprises using seed-specific promoters obtained from flax
XX
XX Claim 14; Fig 3; 68pp; English.
PS
XX
XX The present sequence is a flax 2S storage protein.
CC
XX The present invention relates to a method for expressing non-native genes
CC in flax seeds. The method comprises introducing a chimeric nucleic acid
CC construct containing a seed-specific promoter obtained from flax and a
CC nucleic acid which is non-native to the promoter, into a flax plant cell
CC and growing the plant cell into a mature flax plant capable of setting
CC seed, where the nucleic acid is expressed in the seed under the control
CC of the promoter. The seed-specific promoters obtained from flax are
CC useful for modifying the protein, oil or polysaccharide composition of
CC the flax seeds and seeds of other plant species. The promoters facilitate
CC expression of proteins, including sulphur-rich protein that are found in
CC lupins or Brazil nuts in a seed deficient in sulphurous amino acids,
CC peptides having pharmaceutical value such as anticoagulants, antibodies,
CC vaccines, cytokines, growth factors, interleukins, mammalian proteins,
CC including alpha-1-antitrypsin, anti-obesity proteins, haemoglobin, blood
CC proteins, human serum albumin, insulin, lactoferrin, myoglobin, pulmonary
CC surfactants and proteins of industrial value such as alpha-amylase,
CC arabinase, amyloglucosidase, catalase, cellobiohydrolase, pectinases,
CC phytase, papain and xylanase.

PI Griffin JA, Kallick DA, Tribouley CM, Yue H, Nguyen DB, Tang YT;
PI Lal P, Policky JL, Azimzai Y, Lu DAM, Graul R, Yao MG, Burford N;
PI Hafalia AJA, Baughn MR, Bandman O, Patterson C, Yang J, Xu Y;
PI Gandhi AR, Warren BA, Ding L, Sanjanwala MS, Duggan BM, Lu Y;
XX WPI; 2002-090432/12.
DR N-PSDB; ABK15174.
XX
XX Twelve human receptors (referred to as REPTR-1 to REPTR-12), useful in
PT the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), autoimmune/inflammatory (e.g. osteoarthritis) and cell
PT proliferative (e.g. cancer) disorders -
XX
PS Claim 50; Page 119-121; 157pp; English.
XX
XX This invention relates to twelve human receptors cDNA sequences
CC referred to as REPTR-1 to REPTR-12), and the proteins encoded thereby.
CC The proteins of the invention may have antiinflammatory, cytostatic,
CC immunosuppressive, antiviral, anti-HIV, antiarthritic, muscular active
CC general, anticonvulsant, nootropic, neuroprotective, anti-allergic
CC activities. The sequences of the invention may be used to produce REPTR
CC agonists or antagonists, and the protein sequences may be used to raise
CC anti-REPTR antibodies. These molecules and the REPTR polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal (e.g. gastritis, oesophageal carcinoma,
CC Crohn's disease, irritable bowel syndrome, ulcerative colitis),
CC endocrine (e.g. hypothyroidism, Kallman's disease), autoimmune/
CC inflammatory (e.g. acquired immune deficiency syndrome (AIDS),
CC rheumatoid arthritis, allergies, osteoarthritis, diabetes mellitus,
CC multiple sclerosis, systemic lupus erythematosus), cell proliferative
CC (e.g. cancer), developmental (e.g. Duchenne and Becker muscular
CC dystrophy), neurological (e.g. epilepsy, Alzheimer's disease,
CC Huntington's disease) and reproductive (e.g. infertility, endometriosis)
CC disorders. Numerous other examples of each disorder are given in the
CC specification. The present sequence represents the human REPTR6 protein
XX sequence of the invention.
XX
SQ Sequence 694 AA;

Query Match 9.1%; Score 105; DB 23; Length 694;
Best Local Similarity 23.7%; Pred. No. 0.26;
Matches 51; Conservative 22; Mismatches 72; Indels 70; Gaps 10;
QY 5 AAAAAALCFALVAVACQ-----EVER-QRLRLDQCQEV 41
DB 25 AAAAAKELACQETVPLCKGIGYNTYMPNQFNHDTQDEAGLEVHGFVPLVEIQSPDL 84
QY 42 Q-----ESPLDACRQVLDRLTGGGGGGVGP-----FRWGTGLMRCC 80
DB 85 KFFLCSMYTPICLEDYKPLPPCRSVCR-----AKAGCAPLMRQYGFANPD--RMRC 136
QY 81 QQLQDVSRERC-----AAIRSMVRGYEAMPLEKGMWPGWGRQQQPPPGGGGGGGY 135
DB 137 RLPEQGNPDLTCLMDYNRDLTTAAPSPPRLPPPPG-----EQPPSGSGHGRP--- 185
QY 136 YPCSR-PGEGYGYGGGQGMVPPCRPGTTGGPR 169
DB 186 -FCARPPHRRGGGGGGGDAAPPARGGGGGGKAR 219

RESULT 12
ID AAE34057
XX AAE34057 standard; Protein; 694 AA.
AC AAE34057;
XX
DT 02-MAY-2003 (first entry)
XX
DE FZD 8 protein.
XX
KW Drug screening; toxicology assay; signalling pathway; FZD 8.
XX
OS Unidentified.

XX Key Location/Qualifiers
FH Misc-difference 434
FT /note= "Encoded by ATC"
XX
PN WO2002090992-A2.
XX
PD 14-NOV-2002.
XX
XX 29-APR-2002; 2002WO-CB01946.
XX
XX 04-MAY-2001; 2001GB-0011004.
XX
PA (AXOR-) AXORDIA LTD.
XX
PI Andrews P, Draper J, Walsh J;
XX
DR WPI; 2003-120579/11.
DR N-PSDB; AAD52555.
XX
XX Identifying biologically active agents comprises cloning transfected
PT cells into a cell array, exposing the array to an agent to be tested,
PT and detecting signals generated by a reporter molecule as a result of
PT exposure to the agent -
XX
PS Claim 16; Fig 66; 90pp; English.
XX
XX The present invention relates to a novel screening method which enables
CC the identification of biologically active agents which mediate their
CC effect through the activation of genes. The method involves providing a
CC population of cells stably transfected with a nucleic acid encoding a
CC reporter molecule, cloning the transfected cells into a cell array,
CC exposing the array to at least one agent to be tested and detecting a
CC signal generated by the reporter molecule as a result of exposure to
CC the agent. The method is useful in identifying biologically active agents
CC and the genes through which the agents act, in screening potential drugs
CC for their ability to activate certain drug targets in a high-throughput
CC assay, in identifying relationships between signalling pathways and
CC specific signals that could be useful in eventually directing the
CC differentiation of embryonic stem cells and in toxicology assays by
CC testing for unwanted activation or inhibition of specific signalling
CC pathways. The present sequence is FZD 8 protein used to illustrate the
CC method of the invention.
XX
SQ Sequence 694 AA;
Query Match 9.1%; Score 105; DB 24; Length 694;
Best Local Similarity 23.7%; Pred. No. 0.26;
Matches 51; Conservative 22; Mismatches 72; Indels 70; Gaps 10;
QY 5 AAAAAALCFALVAVACQ-----EVER-QRLRLDQCQEV 41
DB 25 AAAAAKELACQETVPLCKGIGYNTYMPNQFNHDTQDEAGLEVHGFVPLVEIQSPDL 84
QY 42 Q-----ESPLDACRQVLDRLTGGGGGGVGP-----FRWGTGLMRCC 80
DB 85 KFFLCSMYTPICLEDYKPLPPCRSVCR-----AKAGCAPLMRQYGFANPD--RMRC 136
QY 81 QQLQDVSRERC-----AAIRSMVRGYEAMPLEKGMWPGWGRQQQPPPGGGGGGGY 135
DB 137 RLPEQGNPDLTCLMDYNRDLTTAAPSPPRLPPPPG-----EQPPSGSGHGRP--- 185
QY 136 YPCSR-PGEGYGYGGGQGMVPPCRPGTTGGPR 169
DB 186 -FCARPPHRRGGGGGGGDAAPPARGGGGGGKAR 219

RESULT 13
ID ABUS5903
XX ABUS5903 standard; Protein; 694 AA.
AC ABUS5903;
XX

DT 25-MAR-2003 (first entry)
XX Human protein Frizzled-8.
XX Notch; Wnt; embryonic stem cell; embryogenesis; human;
KW differentiation; ligand; Parkinson's disease; Huntington's disease;
KW motor neuron disease; heart disease; diabetes; liver disease;
KW cirrhosis; renal disease; AIDS; acquired immunodeficiency syndrome.
XX Homo sapiens.
XX WO200277204-A2.
PN 03-OCT-2002.
XX 25-MAR-2002; 2002WO-GB011195.
XX 23-MAR-2001; 2001GB-0007296.
PR 23-MAR-2001; 2001GB-0007299.
PR 17-APR-2001; 2001GB-0009346.
XX (AXOR-) AXORDIA LTD.
PA Andrews P, Walsh J, Gokhale P;
PI WPI; 2003-092852/08.
XX N-PSDB; ABX75330.
DR Modulating the differentiation of embryonic stem cells by providing
PT ligands which bind receptors in the Notch and Wnt pathways, useful for
PT treating diseases such as Parkinson's, Huntington's, heart disease,
PT diabetes and AIDS -
XX Disclosure; Fig 72; 121pp; English.
XX The invention relates to modulating the differentiation of an embryonic
CC stem cell, comprising: (a) providing a culture of embryonic stem cells;
CC (b) providing at least one ligand or its active binding fragment,
CC capable of binding its cognate receptor polypeptide expressed by the
CC embryonic stem cell; (c) forming a culture comprising embryonic stem
CC cells and the ligand; and (d) growing the cell culture. Also included
CC are: (i) Modulating the differentiation of embryonic stem cells,
CC comprising: (a) providing a cell transfected with a nucleic acid molecule
CC selected from: (i) any of 9 fully defined Wnt nucleic acid sequences;
CC (ii) a nucleic acid molecule that hybridises to the nucleic acid in
CC (i), and which encodes a ligand capable of modulating embryonic stem
CC cell differentiation, or capable of binding a Wnt receptor; or
CC (iii) nucleic acid molecules which are degenerate as a result of the
CC genetic code to the sequences of (i) or (ii); (b) forming a culture
CC comprising the cell identified in (a) with an embryonic stem cell; and
CC (c) growing the culture for the maintenance and/or differentiation of
CC the embryonic stem cell; (2) Inhibiting the differentiation of embryonic
CC stem cells, comprising: (a) providing at least one polypeptide or its
CC active fragment, that are inhibitors of the Wnt signalling pathway;
CC (b) forming a culture comprising the cell identified in (a) with an
CC embryonic stem cell; and (c) growing the culture for the maintenance of
CC embryonic stem cells in an undifferentiated state; or (3) Inhibiting the
CC differentiation of embryonic stem cells, comprising: (a) providing a cell
CC transfected with a nucleic acid molecule selected from: (i) a molecule
CC encoding a Wnt inhibitory polypeptide; (ii) a molecule which hybridises
CC to the molecule of (i) and encodes a polypeptide capable of inhibiting
CC Wnt signalling; and (iii) nucleic acid molecules which are degenerate as
CC a result of the genetic code to the sequences of (i) or (ii); (b) forming
CC a culture comprising the cell identified in (a) with an embryonic stem
CC cell; and (c) growing the culture for the maintenance of embryonic stem
CC cells in an undifferentiated state; and (4) A cell, therapeutic cell or
CC cell culture obtainable by any of the methods cited above.
CC The therapeutic cell of the present invention is useful in the
CC treatment of an animal, preferably a human, comprising administering a
CC cell composition comprising embryonic stem cells which have been
CC induced to differentiate into at least one cell-type. The cell is also
CC useful for the manufacture of a composition for use in treatment of
CC diseases such as Parkinson's disease, Huntington's disease, motor

CC neuron disease, heart disease, diabetes, liver disease (e.g.
CC cirrhosis), renal disease and AIDS (acquired immunodeficiency syndrome).
CC The present sequence is represents a Wnt or Notch pathway protein
CC (i.e. a ligand for the method of the invention).
XX Sequence 694 AA;
SQ Query Match 9.1%; Score 105; DB 24; Length 694;
Best Local Similarity 23.7%; Pred. No. 0.26;
Matches 51; Conservative 22; Mismatches 72; Indels 70; Gaps 10;
QY 5 AAAAAALCFAALVAVVCOG-----ESPLDACRVLDRQLTGGGGGGVGP-----FRWGTGLRMRC 80
DB 25 AAAAAAKELACQETITVPLCKGIGYNYTMPNQFNHDTQDEAGLEVHFQFWPLVEIQSPDL 84
QY 42 Q-----EPLDACRVLDRQLTGGGGGGVGP-----FRWGTGLRMRC 80
DB 85 KFFLCSMYTPICLEDYKKPLPCRSVCER-----AKAGCAPLMRQYGFAMPD--RMRCD 136
QY 81 QQLQDVSRRCR-----AAIRSMVRGYEAMPPLEKGMWPMGRQOQPPPGGGGGGGY 135
DB 137 RLPEQGNPTLMDYNRDITLTAAPSPRRLLPPPPG-----EQPSSGSGHGRP--- 185
QY 136 YPCSR-PGEGYGGGQGMYPCCRPPTGGPR 169
DB 186 -PGARPPHRRGGGGGGGDAAPPARGGGGGKAR 219
RESULT 14
ABB59015
ID ABB59015 standard; Protein; 348 AA.
XX ABB59015;
AC ABB59015;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 3837.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL03118.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 3837; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 348 AA;

Query Match 9.0%; Score 103.5; DB 22; Length 348;
Best Local Similarity 25.4%; Pred. No. 0.17;
Matches 36; Conservative 11; Mismatches 38; Indels 57; Gaps 4;

QY 48 ACQVLDRLTGCGGGGGVGFPRWGTLRMRCQQQLQDVSRRCRCAAIRSMVRGYEEAMP 107
Db 13 ACNATFLSLGGGGGGGGGSK-----TTYNVIAT 43

QY 108 PLEKGMWPMGRQQPPPGGGGGGGGGYPCSRPG-----EGYGYGGGGGQGMYPPCR 160
Db 44 PSSGGGGGG-----CGGGGGGGHGYSAQGGGGGGHGYAGGHGYGHG----- 85

QY 161 PGTGGGPRIGRVLTKAREYA 182
Db 86 ---HGGSPQIIKVLQEGGGYS 104

RESULT 15
ABB62974
ID ABB62974 standard; Protein; 202 AA.
XX
AC ABB62974;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 15714.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07077.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 15714; 21pp + Sequence Listing; English.
XX

The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
XX
The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 202 AA;

Query Match 9.0%; Score 103; DB 22; Length 202;
Best Local Similarity 28.3%; Pred. No. 0.1;
Matches 34; Conservative 12; Mismatches 48; Indels 26; Gaps 4;

QY 60 GGGGGGVPFRWGTGLRMRCQQQLQDVSRRCRCAAIRSMVRGYEEAMPPL 114
Db 72 GGGGGGGVGGVGGYGGPAPLTEQVGYGGTQ---GGYGAQKGYGASVPPNEFALGSGQG 128

QY 115 PMGRQQQPPPGG---GGGGGGYYYPCSRPGEGYGYGGGQGMYPPCRPGTTGGGPRI 170
Db 129 RLGGYGNATPQSGKLGYGGGGGYRRPAPQSQLYAAGE-----LGAGPQL 174

Search completed: November 29, 2003, 17:05:01
Job time : 62 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2003, 16:58:04 ; Search time 26 Seconds
(without alignments)
761.952 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAAALCFALVAV.....MMCRLEPQECISFGSDQY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96169682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	439	38.2	186	JC4784	alpha-globulin pre
2	435	37.9	186	WMR219	19K globulin precu
3	289.5	25.2	650	A24266	glutenin high mole
4	287	25.0	705	S18733	glutenin high mole
5	282.5	24.6	707	S52390	D-hordein precu
6	281	24.5	648	S04832	glutenin high mole
7	157	13.7	830	S15720	glutenin high mole
8	156	13.6	791	JN0690	glutenin, high-mol
9	156	13.6	815	S30843	glutenin high mole
10	155.5	13.5	848	S02262	glutenin high mole
11	152	13.2	815	JN0689	glutenin high mole
12	151.5	13.2	789	A30843	glutenin, high-mol
13	151.5	13.2	838	1 EEWTHW	glutenin, high mol
14	147.5	12.8	733	JC2099	glutenin, high mol
15	133.5	11.6	161	JC4966	high-molecular-we
16	126	11.0	323	S38887	2S albumin - commo
17	117.5	10.2	309	S10889	proline-rich prote
18	114.5	10.0	146	S14946	2S seed storage pr
19	114	9.9	297	T06500	alpha/beta-gliadin
20	112.5	9.8	258	1 RZCS	2S seed storage pr
21	112	9.8	295	S01062	2S seed storage pr
22	111.5	9.7	307	S10015	alpha/beta-gliadin
23	111	9.7	296	S07361	alpha/beta-gliadin
24	110	9.6	101	1 EEWTL	glutenin 1 - wheat
25	110	9.6	139	2 T09878	albumin 2S storage
26	110	9.6	145	2 S13376	CM2 protein - duru
27	110	9.6	326	2 A41732	heterogeneous ribo
28	110	9.6	386	1 S22315	snRNP-associated p
29	109.5	9.5	291	2 T06498	alpha/beta-gliadin

allergen RA14C pre
seed allergen RAG2
alpha/beta-gliadin
2S albumin - Brazi
late embryogenesis
eggshell protein p
hypothetical prote
alpha-amylase inhi
alpha/beta-gliadin
proline-rich prote
2S albumin precurs
trypsin/factor XII
probable nuclear a
proline-rich prote
albumin 2S storage
trypsin inhibitor

ALIGNMENTS

RESULT 1

JC4784

alpha-globulin precursor - rice

C:Species: Oryza sativa (rice)

C:Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 20-Jun-2000

C:Accession: JC4784

R:Nakase, M.; Hotta, H.; Adachi, T.; Aoki, N.; Nakamura, R.; Masumura, T.; Tanaka, K.; Me

Gene 170, 223-226, 1996

A:Title: Cloning of the rice seed alpha-globulin-encoding gene: Sequence similarity of t

A:Reference number: JC4784; MUID:96235139; PMID:8666249

A:Accession: JC4784

A:Molecule type: DNA

A:Residues: 1-186 <NAK>

A:Cross-references: DDBJ:P50643; NID:G840704; PIDN:BAA09308.1; PID:G1783206

A:Experimental source: seed

C:Genetics:

A:Gene: Glb

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: Globulin; seed

F;1-23/Domain: signal sequence #status predicted <SIG>

F;24-186/Product: alpha-globulin #status predicted <MAT>

Query Match 38.2%; Score 439; DB 2: Length 186;
Best Local Similarity 51.4%; Pred. No. 1.6e-29;
Matches 108; Conservative 20; Mismatches 54; Indels 28; Gaps 9;
QY 1 MAKIAAAAAAALCFALVAVAVCQGEVERQRLRLDQCWQEVQESPLDQCRQVLDRLQITGG 60
DB 1 MASKVVFVFAAL-MAAMVAISGAQLSESEMRFRDRCQREVQDPSPLDQCRQVLDRLQITGR 59
QY 61 GGGGGVGPFRMTGLMRCCQQLQDVSRRCRCACATRSMTVRGYEEAMP-FLKQGWPMWGRQ 119
DB 60 ERFQPMFRFPFGALGLMRCCQQLQDVSRRCRCACATRSMTVRGYEEAMP-FLKQGWPMWGRQ 119
QY 120 QPPPPGGGGGGGGYYPGSRPGEYGYGQGGORQMY---PPCRPGTTGGPRIGVRILT 176
DB 120 YY-----GGEG-----SSSEQGY-YGEGSEEGYGEQQQPGMT-----RVRLT 158
QY 177 KAREYAAGLPMWCRLEPQECISFGSDQY 206
DB 159 RARQYAAQLPMCRV-EPOQCSIFAAG-QY 186

RESULT 2

WMR219

19K globulin precursor - rice

N;Alternate names: alpha-globulin

C:Species: Oryza sativa (rice)

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 18-Jun-1999

C:Accession: S20024; S25735; PN0497

R;Shorrosh, B.S.; Wen, L.; Zen, K.C.; Huang, J.K.; Pan, J.S.; Hermodson, M.A.; Tanaka, K.

Qy	1	MAKIAAAAAALCFAALVAVAVCQGVVERQRLRDLQCWQEVQESPLDACRQVLDROLTGG	60
Db	1	MAKRLVLAFAV--IALVALTAAEGASRQ-----LQCRELQESSLEACRQVVVDQQL---	51
Qy	61	GGGGGVPFRWGTGLRMRCQQLQDVSRRCRCAAIRSMVRGYEE-AMPPLEKGMWP----	115
Db	52	-----AGRLPWSLGLQMRCCQLRDVSAKRSVAVSQVAREQTVVPPKGGSFYPGETT	106
Qy	116	-----WG-----RQ-----QPPPPQGGGGGGGG-----	133
Db	107	PLQQLQGLFWGTSSQTVOGYPSVTSFRQGSYYPQASFPQPGQCGQPGKWOEPGCGQ	166
Qy	134	YYPYCS--RPGEYGYGQGGQRQMYPPCRPTGTGGQPRIGR	172
Db	167	WYPTSLQPGQGGQGIKGG--KGGYPTSLQPGQGGQGIQ	205

RESULT 4
 S18733
 glutenin high molecular weight chain 1B9 precursor - wheat
 C;Species: Triticum aestivum (common wheat)
 C;Date: 08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
 C;Accession: S18733
 R;Halford, N.G.; Forde, J.; Anderson, O.D.; Greene, F.C.; Shewry, P.R.
 Theor. Appl. Genet. 75, 117-126, 1987
 A;Title: The nucleotide and deduced amino acid sequences of an HMW glutenin subunit A and 1D.
 A;Reference number: S18733
 A;Accession: S18733
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-705 <HAL>
 A;Cross-references: EMBL:X61026; NID:g22089; PIDN:CAA43361.1; PID:g22090
 C;Superfamily: glutenin

Query Match	25.0%;	Score	287;	DB	2;	Length	705;
Best Local Similarity	32.9%;	Pred. No.	2.3e-16;				
Matches	85;	Conservative	24;	Mismatches	47;	Indels	102;
						Gaps	13;

Qy	17	LVAVAVCQGVVERQRLRDLQCWQEVQESPLDACRQVLDROLTGGGGGGVGPFRWGTGLR	76
Db	15	LVALTAEGASRQ-----LQCRELQESSLEACRQVVVDQQL-----AGRLPWSLGLQ	62
Qy	77	MRCQQLQDVSRRCRCAAIRSMVRGYEE-AMPPLEKGMWP-----WGR----	118
Db	63	MRCQQLRDVSAKCRPVAVSQVVRQEQTVVPPKGGSFYPGETTPLQQLQGVIFWGTSSQ	122
Qy	119	-----QQP-----PPGG-----GGGGGGY-----	134
Db	123	TVQGYPSVSSPQGGYYPQASPPQGGQPGKQWELGQGGYPTSLHSGGGQCG	182
Qy	135	YYPYCS--RPGEYGYGQGGQGGQGGYPTSLQPGQGGQGGYPTSLQPGQGGQGGYPTSPHQRQPGQG	242
Db	183	YPPSSLQPGQGGQGGQGGYPTSLQPGQGGQGGYPTSLQPGQGGQGGYPTSPHQRQPGQG	242
Qy	168	PRIGR-VRLTKAREVAAAG	184
Db	243	QVIGQGGQGLGGRQIGQG	260

RESULT 5
 S52390
 D-hordein precursor - barley
 C;Species: Hordeum vulgare (barley)
 C;Date: 08-May-1995 #sequence_revision 26-Jul-1996 #text_change 20-Aug-1999
 C;Accession: S52390; S23921
 R;Sorensen, M.B.; Muller, M.; Simpson, D.
 submitted to the EMBL Data Library, February 1995
 A;Description: Hordein promoter methylation and transcriptional activity in wild type.
 A;Reference number: S52390
 A;Accession: S52390
 A;Molecule type: DNA

A;Residues: 1-475 <SOR>
 A;Cross-references: EMBL:X84368; NID:g671536; PIDN:CAA59104.1; PID:g671537
 R;Halford, N.G.; Tatham, A.S.; Sui, E.; Daxoda, L.; Dreyer, T.; Shewry, P.R.
 Biochim. Biophys. Acta 1122, 118-122, 1992
 A;Title: Identification of a novel beta-turn-rich repeat motif in the D hordeins of barley
 A;Reference number: S23921; MUID:92353095; PMID:1643086
 A;Accession: S23921
 A;Molecule type: mRNA
 A;Residues: 267-355, 'P', 357-359, 'Q', 361-458, 'Y', 460-707 <HAL>
 A;Cross-references: EMBL:X68072; NID:g18969; PIDN:CAA48209.1; PID:g18970
 C;Genetics:
 A;Gene: hor3
 C;Superfamily: glutenin
 C;Keywords: seed; storage protein
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-707/Product: D-hordein #status predicted <MAT>

Query Match 24.6%; Score 282.5; DB 2; Length 707;
 Best Local Similarity 36.2%; Pred. No. 5.5e-16;
 Matches 79; Conservative 22; Mismatches 52; Indels 65; Gaps 11;

QY 1 MAKIAAAAAALCFALVAVAVCOGEVE-----RORLRDLCQWQVEVQSPDLACRQVLD 54
 DB 1 MAKRLVLFVAVI--VALVALTTAEREINGNNIFLDSRQLQCERELQSSLEACRRVVD 58
 QY 55 RQLTGGGGGGVGFWRGTLRMRCQQQLQDVSRRCRAIRSMVRGYEE----- 104
 DB 59 QQL-----VGQLPWSGTGLQWCCQLRDVSPCRPVALSQVRYEQTEVPKSGS 110
 QY 105 -----AMPLEK-GWV-----PWGRQQPPPGGGGGGGGGY-----YPC 138
 DB 111 FYPGGTAPPLQGGWGTSGKNNYPDTSSQSQWQGGVQHSQTPGQGGQGGSYPG 170
 QY 139 S-----RPGGYGVGCGGRQ--MYP-----PCRGTGG 166
 DB 171 STFPQPGGQ-----QPGQRQPSYSPATFPQPGGQGG 205

RESULT 6
 S04832
 glutenin high molecular weight chain (Dy10) - wheat
 C;Species: Triticum aestivum (common wheat)
 C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 20-Aug-1999
 C;Accession: S04832; S06645
 R;Anderson, O.D.; Greene, F.C.; Yip, R.E.; Halford, N.G.; Shewry, P.R.; Malpica-Romero,
 Nucleic Acids Res. 17, 461-462, 1989
 A;Title: Nucleotide sequences of the two high-molecular-weight glutenin genes from the D
 A;Reference number: S02262; MUID:89098419; PMID:2563152
 A;Accession: S04832
 A;Status: not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-648 <AND>
 A;Cross-references: GB:X12929; NID:gl360617; PIDN:CAA31396.1; PID:g21751
 R;Goldsbrough, A.P.; Bulleid, N.J.; Freedman, R.B.; Flavell, R.B.
 Biochem. J. 263, 837-842, 1989
 A;Title: Conformational differences between two wheat (Triticum aestivum) 'high-molecular
 A;Reference number: S06644; MUID:90088430; PMID:2597130
 A;Accession: S06645
 A;Status: not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 126-474, 'QG', 477-648 <GOL>
 C;Genetics:
 A;Gene: Glu-D1-2b
 C;Superfamily: glutenin

Query Match 24.5%; Score 281; DB 2; Length 648;
 Best Local Similarity 36.6%; Pred. No. 6.8e-16;
 Matches 78; Conservative 23; Mismatches 46; Indels 66; Gaps 11;

QY 1 MAKIAAAAAALCFALVAVAVCOGEVEVQRLDLCQWQVEVQSPDLACRQVLDRLTGG 60
 DB 1 MAKRLVLFVAVV--IALVALTTAEGEASQ-----LQCERELQSSLEACRRVVDQQL--- 51

QY 61 GGGGGVGFPRWGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE-AMPLEKGWMP----- 115
 DB 52 -----AGRLPWSGTGLQMRCCQLRDVSAKCRSVAVSQARQYEQTVVPPKGGSYFGETT 106
 QY 116 -----WG-----RQ-----QPPPPGGGGGGGG- 133
 DB 107 PLQQLQGGIFWGTSTQTVGGYYPGVTSRQSGSYVPGQASPPQPGQQQPGKWKQEPGQGGQ 166
 QY 134 YYPGCS--RPGEGYGVGCGGQGRQMYPPC--RPG 162
 DB 167 WYPTSLQPGGGGQGGQIGK-GQGGYPTSLQPG 198

RESULT 7
 S15720
 glutenin high molecular weight chain 1A1 precursor - wheat
 C;Species: Triticum aestivum (common wheat)
 C;Date: 08-Jun-1994 #sequence_revision 12-May-1995 #text_change 20-Aug-1999
 C;Accession: S15720
 R;Halford, N.G.; Field, J.M.; Blair, H.; Urwin, P.; Moore, K.; Robert, L.; Thompson, R.;
 submitted to the EMBL Data Library, July 1991
 A;Description: Analysis of HMW glutenin subunit encoded by chromosome 1A of bread wheat:
 A;Reference number: S15720
 A;Accession: S15720
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-830 <HAL>
 A;Cross-references: EMBL:X61009; NID:g21742; PIDN:CAA43331.1; PID:g21743
 C;Superfamily: glutenin

Query Match 13.7%; Score 157; DB 2; Length 830;
 Best Local Similarity 24.6%; Pred. No. 1.7e-05;
 Matches 63; Conservative 17; Mismatches 42; Indels 134; Gaps 11;

QY 11 ALCFALVAVAVCOGEVEVQRLDLCQWQVEVQSPDLACRQVLDRLTGGGGGGVGPFR 70
 DB 9 AAVVALVALTAEGEASQ-----LQCERELQSSLEACRRVVD----- 48
 QY 71 WGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE-AMPLEKGWMP----- 115
 DB 49 -----QQLRDVSPGCVGGGFVARQYEQTVVPPKGGSYFGETTTPQQLQSSI 98
 QY 116 -W-----GQQQP-----PQGG----- 127
 DB 99 LWGIPALLRRYLVTSPPQVSYVPCQASSRPGGQQPGQGGQVEYLLSPQSGQWQPP 158
 QY 128 GGGGGY-----YPPCS-----RPGGYGVGG----- 150
 DB 159 GQGAQGYPTSPQSGQEQGYPTSPWQPEQLQPTQGGQQPQGGQQLRGGQGGQS 218
 QY 151 --GQRQMYPPC--RPG 162
 DB 219 GQGPRIYPTSSQPG 234

RESULT 8
JN0690

glutenin, high-molecular-weight Bx7 chain precursor - wheat

C;Species: Triticum aestivum (common wheat)

C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 14-Jul-1994

C;Accession: JN0690

R;Anderson, O.D.; Greene, F.C.

Theor. Appl. Genet. 77, 689-700, 1989

A;Title: The characterization and comparative analysis of high-molecular-weight glutenin

A;Reference number: JN0689

A;Accession: JN0690

A;Molecule type: DNA

A;Residues: 1-791 <AND>

C;Comment: The main wheat storage proteins are divided into two groups. The glutenins, c

A families.

C;Superfamily: glutenin

C;Keywords: seed; storage protein

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-791/Product: glutenin, HMW Bx7 chain #status predicted <MAT>

Query Match 13.6%; Score 156; DB 2; Length 791;
Best Local Similarity 28.1%; Pred. No. 2e-05;
Matches 59; Conservative 19; Mismatches 54; Indels 78; Gaps 10;
QY 1 MAKIAAAAAALCFALVAVAVCGEVEVERQLRDLCQWQVQESPLDACRQVLDRLQTGG 60
DB 1 MAKRLVLFVAVV--VALVALTAAGEASGQ---LQC-----EHELEACQVVD----- 43
QY 61 GGGGGVGPFRWGTGLRMRCQQQLQDVSRRCRAIRSMVRYEE--AMPPEKXGWP--- 115
DB 44 -----QQLRDVSPGCRPTVSPGTRQYEQQPVWFSKAGSFYPSSET 83
QY 116 -----WG-----RQQP-----PPCGGGGGGGYYPYPCSRPGE 143
DB 84 TPSQQLQWIFWGIPLLALRYVSVTSQQSGSYPGQASQQSGGQGPQGE---QQPGQ 140
QY 144 GYGYGGGGOROM-YPCRCPTGCGPRIGR 172
DB 141 GQHQQPGRQQGYPTSPQPGQGGQQLGQ 170

RESULT 9

B30843
glutenin high molecular weight chain Ax2 precursor - wheat

C;Species: Triticum aestivum (common wheat)
C;Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 15-Nov-1996
C;Accession: B30843
R;Anderson, O.D.; Greene, F.C.
submitted to GenBank, January 1989
A;Reference number: A94515
A;Accession: B30843
A;Molecule type: DNA
A;Residues: 1-815 <AND>
C;Superfamily: Glutenin
C;Keywords: seed; storage protein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-815/Product: glutenin Ax2 chain #status predicted <MAT>
F;108-773/Region: glutamine/glycine/proline-rich

Query Match 13.6%; Score 156; DB 2; Length 815;
Best Local Similarity 24.8%; Pred. No. 2.1e-05;
Matches 62; Conservative 17; Mismatches 43; Indels 128; Gaps 11;
QY 11 ALCFALVAVAVCGEVEVERQLRDLCQWQVQESPLDACRQVLDRLQTGGGGGGVGPFR 70
DB 9 AAVVALVALTAAGEASGQ---LQCERELQEHSLKACRQVVD----- 48
QY 71 WGTGLMRCCQQLQDVSRRCRAIRSMVRYEE--AMPPEKXGWP----- 115
DB 49 -----QQLRDVSPGCRPTVSPGTRQYEQQPVWFSKAGSFYPSSET 98
QY 116 -WG-----RQQP-----PPCGG-----GGGGGG 133
DB 99 LMGIPALLRRYLYSVTSPOQVSYYPGQASQRPQGGQYLYTSPQSGWQPGQGG 158
QY 134 Y-----YPPCS-----RPGEGYGGG-----GQRQ 154
DB 159 YPTSPQSGQKQPGYYPYTPSPWPEQLQQTQGGQRPQGGQQLRQGGGQGGQGP 218
QY 155 MYPPC--RPG 162
DB 219 YPTSSQQPG 228

RESULT 10

S02262

glutenin high molecular weight chain Dx5 - wheat

C;Species: Triticum aestivum (common wheat)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 03-Feb-1994
C;Accession: S02262
R;Anderson, O.D.; Greene, F.C.; Yip, R.E.; Halford, N.G.; Shewry, P.R.; Malpica-Romero,

Nucleic Acids Res. 17, 461-462, 1989

A;Title: Nucleotide sequences of the two high-molecular-weight glutenin genes from the D-

A;Reference number: S02262; MUID:89098419; PMID:2563152

A;Accession: S02262

A;Molecule type: DNA

A;Residues: 1-848 <AND>

A;Cross-references: EMBL:X12928

C;Genetics:

A;Gene: Glu-D1-1b

C;Superfamily: glutenin

Query Match 13.5%; Score 155.5; DB 2; Length 848;
Best Local Similarity 26.0%; Pred. No. 2.3e-05;
Matches 61; Conservative 17; Mismatches 52; Indels 105; Gaps 10;

QY 1 MAKIAAAAAALCFALVAVAVCGEVEVERQLRDLCQWQVQESPLDACRQVLDRLQ 57
DB 1 MAKRLVLFVAVV--VALVALTAAGEASGQ---LQCERELQEHSLKACRQVVD--- 51
QY 58 TGGGGGGVGPFRWGTGLRMRCQQQLQDVSRRCRAIRSMVRYEE--AMPPEKXGWP 115
DB 52 -----QQLRDVSPGCRPTVSPGTRQYEQQPVWFSKAGSFYPSSET 88
QY 116 -----WG-----RQQP-----GGGGGG 134
DB 89 GETTPQQLQWIFWGIPLLALRYVSVTSPOQVSYYPGQASQRPQGGQYLYTSPQSGWQPGQGG 148
QY 135 -----YPPCS-----RPGEGYGGGGYYPYPCSRPGE 162
DB 149 PTPSPQSGQKQPGYYPYTPSPWPEQLQQTQGGQRPQGGQQLRQGGGQGGQGP 203

RESULT 11

JN0689

glutenin, high-molecular-weight Ax2* chain precursor - wheat

C;Species: Triticum aestivum (common wheat)

C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 15-Nov-1996

C;Accession: JN0689

R;Anderson, O.D.; Greene, F.C.

Theor. Appl. Genet. 77, 689-700, 1989

A;Title: The characterization and comparative analysis of high-molecular-weight glutenin

A;Reference number: JN0689

A;Accession: JN0689

A;Molecule type: DNA

A;Residues: 1-815 <AND>

A;Note: The authors translated the codon CTA for residue 11 as Val, CAT for residue 496 as a families.
C;Comment: The main wheat storage proteins are divided into two groups. The glutenins, c

C;Genetics:

A;Gene: Ax2*

C;Superfamily: glutenin

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-815/Product: glutenin, high-molecular-weight Ax2* chain #status predicted <MAT>

Query Match 13.2%; Score 152; DB 2; Length 815;
Best Local Similarity 24.4%; Pred. No. 4.4e-05;
Matches 61; Conservative 18; Mismatches 43; Indels 128; Gaps 11;

QY 11 ALCFALVAVAVCGEVEVERQLRDLCQWQVQESPLDACRQVLDRLQTGGGGGGVGPFR 70
DB 9 AAVVALVALTAAGEASGQ---LQCERELQEHSLKACRQVVD----- 48
QY 71 WGTGLMRCCQQLQDVSRRCRAIRSMVRYEE--AMPPEKXGWP----- 115
DB 49 -----QQLRDVSPGCRPTVSPGTRQYEQQPVWFSKAGSFYPSSET 98
QY 116 -WG-----RQQP-----PPCGG-----GGGGGG 133
DB 99 LMGIPALLRRYLYSVTSPOQVSYYPGQASQRPQGGQYLYTSPQSGWQPGQGG 158
QY 134 Y-----YPPCS-----RPGEGYGGG-----GQRQ 154
DB 159 YPTSPQSGQKQPGYYPYTPSPWPEQLQQTQGGQRPQGGQQLRQGGGQGGQGP 218

QY 155 MYPPC--RPG 162
Db 219 YPPTSQQPG 228

RESULT 12
A30843
glutenin high molecular weight chain Bx7 precursor - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 20-Aug-1999
C:Accession: A30843
R:Anderson, O.D.; Greene, F.C.
submitted to GenBank, January 1989
A:Reference number: A94515
A:Accession: A30843
A:Molecule type: DNA
A:Residues: 1-789 <AND>
A:Cross-references: EMBL:M22209; NID:g170744; PIDN:AAA34291.1; PID:g170745
C:Superfamily: glutenin
C:Keywords: seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-789/Product: glutenin Bx7 chain #status predicted <MAT>
F:103-747/Region: glutamine/glycine/proline-rich

Query Match 13.2%; Score 151.5; DB 2; Length 789;
Best Local Similarity 24.4%; Pred. No. 4.7e-05;
Matches 62; Conservative 21; Mismatches 52; Indels 119; Gaps 9;

QY 1 MAKIAAAAAALCFALVAVAVCOGVEVERQLRDLCQWQEVQESPLDACRQVLDRLTGG 60
Db 1 MAKRLVLFVAVV--VALVALTVAEGEASQ-----LQC-----EHELEACQVVD----- 43

QY 61 GGGGGVGFPRWGTLRMRCQQQLQDVSRRCRAAIRSMVRGYEE----- 104
Db 44 -----QQLRDVSPGCRPITVSPGTRQYEQQPVVASKAGSFYPTSP 83

QY 105 -----AMPPLEKGMWPW-----GROQOP----- 122
Db 84 TFSQQLQQLMFWGIPALLRRYPSVTSSQGSYYPGQASPPQSGGQGGQPGQGGQ 143

QY 123 -----PPQGG-----GGGQGGYYPCSRPGEVGYGQGGQ-----RQMYPP 158
Db 144 DQPGQORQGGYPTSPQPGQGGQGGQGGQGGYPTSPQPGQKQAGQGGQGGQGGYYP 203

QY 159 CRPGTTGGGPRIGR 172
Db 204 TSPQSGGQGGQPGQ 217

RESULT 13
EETHW
glutenin, high molecular weight chain precursor - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C:Accession: A24107
R:Sugiyama, T.; Rafalski, A.; Peterson, D.; Soll, D.
Nucleic Acids Res. 13, 8729-8737, 1985.
A:Title: A wheat HMW glutenin subunit gene reveals a highly repeated structure.
A:Reference number: A24107; MUID:86093674; PMID:3001648
A:Accession: A24107
A:Molecule type: DNA
A:Residues: 1-838 <SUG>
A:Cross-references: GB:X03346; NID:g21784; PIDN:CAA27052.1; PID:g736319
A:Experimental source: cv. Yanhill
C:Comment: Glutenins, like gliadins, are high in glutamine and proline but differ in con
C:Superfamily: glutenin
C:Keywords: seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-838/Product: glutenin, HMW chain #status predicted <MAT>

Query Match 13.2%; Score 151.5; DB 1; Length 838;
Best Local Similarity 28.7%; Pred. No. 5e-05;

Matches 58; Conservative 18; Mismatches 45; Indels 81; Gaps 10;

QY 1 MAKIAAAAAALCFALVAVAVCOGVEVERQLRDLCQWQEVQESPLDACRQVLDRLQ 57
Db 1 MAKRLVLFVAVV--VALVALTVAEGEASQ-----LQCERELQELQERELXACQVWD--- 51

QY 58 TGGGGGGVGFPRWGTLRMRCQQQLQDVSRRCRAAIRSMVRGYEE----- 104
Db 52 -----QQLRDVSPGCRPITVSPGTRQYEQQPVVASKAGSFYPTSP 88

QY 105 -----AMPPLEKGMWPW-----GROQOP----- 122
Db 89 ETTPPQQLQORIFWGIPALLRRYPSVTSPQGSYYPGQASPPQSGGQGGQGGYYP 142

QY 143 EGVGYGQGGQRMYP--PCRPG 162
Db 141 QSQSGQ--GQGGYPTSPQPG 161

RESULT 14
JC2099
glutenin, high molecular weight chain Bx17 - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 23-Mar-1995
C:Accession: JC2099; JQ2261
R:Reddy, P.; Appels, R.
Theor. Appl. Genet. 85, 616-624, 1993
A:Title: Analysis of a genomic DNA segment carrying the wheat high-molecular-weight
A:Reference number: JC2099
A:Accession: JC2099
A:Molecule type: DNA
A:Residues: 1-753 <RED>
A:Experimental source: leaf
A:Genetics:
A:Gene: Bx17
A:Map position: 1B
C:Superfamily: glutenin
C:Keywords: seed; storage protein

Query Match 12.8%; Score 147.5; DB 2; Length 753;
Best Local Similarity 23.8%; Pred. No. 9.7e-05;
Matches 58; Conservative 20; Mismatches 49; Indels 117; Gaps 8;

QY 11 ALCFALVAVAVCOGVEVERQLRDLCQWQEVQESPLDACRQVLDRLTGGGGGGVGPFR 70
Db 9 AAVVALVALTVAEGEASQ-----LQC-----EHELEACQVVD----- 43

QY 71 WGTGLMRCCQQLQDVSRRCRAAIRSMVRGYEE----- 104
Db 44 -----QQLRDVSPGCRPITVSPGTRQYEQQPVVASKAGSFYPTSP 93

QY 105 -----AMPPLEKGMWPW-----GROQOP----- 122
Db 94 FWGIPALLRRYPSVTSSQGSYYPGQASPPQSGGQGGQGGQGGQGGQGGQ 153

QY 123 -----PPQGG-----GGGQGGYYPCSRPGEVGYGQGGQ-----RQMYPPCRGTGGGP 168
Db 154 YPPTSPPQGGQGGQGGQGGYPTSPQPGQKQAGQGGQGGQGGQGGYPTSP 213

QY 169 RIGR 172
Db 214 QPGQ 217

RESULT 15
JC4966
high-molecular-weight glutenin - wheat
N:Alternate names: prolamine
C:Species: Triticum aestivum (common wheat)
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 09-May-1997
C:Accession: JC4966
R:Anderson, O.D.; Kuhl, J.C.; Tam, A.
Gene 174, 51-58, 1996

A;Title: Construction and expression of a synthetic wheat storage protein gene.
A;Reference number: JC4966; MUID:97017127; PMID:8863728
A;Accession: JC4966
A;Molecule type: DNA
A;Residues: 1-161 <AND>
A;Note: the authors translated the codon GAT for residue 12 and GAC for residue 31 as Glu
C;Comment: This protein belongs to one class of wheat prolamines.
C;Superfamily: glutenin

Query Match 11.6%; Score 133.5; DB 2; Length 161;
Best Local Similarity 27.3%; Pred. No. 0.00035;
Matches 54; Conservative 19; Mismatches 64; Indels 61; Gaps 9;

QY 24 QGVEVERQLRDLCQ---WQEVQSSPLDACRQVLDRLQLTGGGGGGVGGPPFRWGTGLRMRC 80
Db 2 EGASEQ-----LQCDRELQELQRELKACQVMD----- 31

QY 81 QQLQDVSRRCRCAAIRSMVRGYEE--AMPPELEKGMWPMGRQQQPPQ-----GGGGGQG 132
Db 32 QQLRDISPECHPVVVSFVAGYEQIIVVP--KGGTFYPGETTPPQQLQRIFWGIPALL 89

QY 133 GYYPGCSRPGEGYGGGGGQRMYPGCRPCTTGGGPRI-----GRVLTAKREYAAGLP 186
Db 90 KRYTPSVTCPPQVSYPG---QASPORSRDITSSSYHVSVEHQASLKVAKAQQLAAQLP 146

QY 187 MMCRLSEPPQECISIFSGD 204
Db 147 AMCRL-----EGGD 155

Search completed: November 29, 2003, 17:07:05
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2003, 15:54:39 ; Search time 18 Seconds
(without alignments)
538.195 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAAALCPAALVAV.....MMCRLEPQECISFGSDQY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Description
1	435	37.9	P29835 Oryza sativ
2	289.5	25.2	P08488 triticum ae
3	281	24.5	P10387 triticum ae
4	156.5	13.6	P10387 triticum ae
5	151.5	13.2	P08488 triticum ae
6	114.5	10.0	P04403 bertholleti
7	114	9.9	P04724 triticum ae
8	112.5	9.8	P01089 ricinus com
9	112	9.8	P15461 helianthus
10	111.5	9.7	P18573 triticum ae
11	111	9.7	P04726 triticum ae
12	110	9.6	P02861 triticum ae
13	110	9.6	P16851 triticum ae
14	110	9.6	P48910 drosophila
15	109.5	9.5	P04722 triticum ae
16	109	9.5	Q01885 oryza sativ
17	106.5	9.3	P16850 triticum ae
18	106.5	9.3	P04727 triticum ae
19	105	9.1	P01088 zea mays (m
20	105	9.1	Q9H461 homo sapien
21	105	9.1	P33485 pseudorabie
22	103	9.0	Q61545 mus musculu
23	102.5	8.9	P23463 mus musculu
24	102	8.9	P01087 eleusine co
25	102	8.9	P10846 hordeum vul
26	102	8.9	P08453 triticum ae
27	101.5	8.8	Q63796 rattus norv
28	101	8.8	P10979 zea mays (m
29	101	8.8	P35085 dictyosteli
30	101	8.8	Q12906 h interieuk
31	100.5	8.8	P25074 oryza sativ
32	100.5	8.8	Q9bzw0 homo sapien
33	100.5	8.8	Q01860 homo sapien

34	100	8.7	304	1	GDB1_WHEAT	P04729 triticum ae
35	100	8.7	465	1	GRP2_PHAVU	P10496 phaseolus v
36	100	8.7	656	1	ERS_HUMAN	Q01844 homo sapien
37	99.5	8.7	145	1	IAA_HORVU	P28041 hordeum vul
38	99.5	8.7	212	1	EGG1_SCHJA	P19470 schistosoma
39	99.5	8.7	220	1	AVE3_AVEA	P80356 avena sativ
40	98.5	8.6	124	1	IAA1_WHEAT	P01085 triticum ae
41	98.5	8.6	289	1	HOG3_HORVU	P80198 hordeum vul
42	97.5	8.5	186	1	GDA8_WHEAT	P04728 triticum ae
43	97	8.4	169	1	GRP2_SINAL	P49311 sinapis alb
44	97	8.4	461	1	STR3_RAT	P58405 rattus norv
45	96.5	8.4	512	1	ANX7_XENLA	Q92125 xenopus lae

ALIGNMENTS

RESULT 1
GL19_ORYSA
ID GL19_ORYSA STANDARD; PRT; 186 AA.

AC P29835;
DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE 19 kDa globulin precursor (Alpha-globulin).

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=cv. Nipponbare; TISSUE=Endosperm;

RX MEDLINE=92119226; PubMed=1731968;

RA Shorrosh B.S., Wen L., Zen K.C., Huang J.K., Pan J.S., Hermanson M.A.,

RA Tanaka K., Muthukrishnan S., Rebeck G.R.;

RT "A novel cereal storage protein: molecular genetics of the 19 kDa

RT globulin of rice.";

RL Plant Mol. Biol. 18:151-154(1992).

RN [2]

RP SEQUENCE OF 5-186 FROM N.A.

RC STRAIN=cv. Lamont; TISSUE=Endosperm;

RX MEDLINE=93277591; PubMed=8503935;

RA Krishnan H.B., Pueppke S.G.;

RT "Nucleotide sequence of an abundant rice seed globulin: homology with

RT the high molecular weight glutelins of wheat, rye and triticale.";

RL Biochem. Biophys. Res. Commun. 193:460-466(1993).

CC -!- FUNCTION: SEED STORAGE PROTEIN.

CC -!- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X63990; CAA45400.1; -

DR EMBL; L12252; AAA72362.1; ALT_INIT.

DR PIR; S20024; WNR219.

DR Gramene; P29835; -

DR InterPro; IPR003612; AAI.

DR InterPro; IPR001419; Glutenin.

DR Pfam; PF00234; tryp_alpha_amyl; 1.

DR PRINTS; PR00210; GLUTENIN.

DR SMART; SM00499; AAI; 1.

KW Signal; Seed storage protein. POTENTIAL.

FT SIGNAL 1 22

FT CHAIN 23 186 19 kDa GLOBULIN.

SQ SEQUENCE 186 AA; 21050 MW; 9E09BA74CB0B6810 CRC64;

Query Match 37.9%; Score 435; DB 1; Length 186;

Query Match	25.2%;	Score	289.5;	DB 1;	Length	660;			
Best Local Similarity	36.2%;	Pred. No.	7.8e-16;						
Matches	80;	Conservative	24;	Mismatches	52;	Indels	65;	Gaps	10;
QY	1	MAKIAAAAAAAAAALCF	AAALVAVAVACQGE	VEVERQRLDLQC	MQEVSPLDAC	RQVLDRQLTGG	60		
DB	1	MAKRLVLPAAVV--	IALVALTTAE	GAASRQ----	LQERELOQESSLE	ACRQVVQQQL---	51		
QY	61	GGGGVGPPFRWGTGL	RMRCQQQLQDVRS	ECRCAATRSVMV	GYEE--	AMPLEKGWNP----	115		
DB	52	-----AGRLPWSTGL	QMRCQQQLRDVSA	KCRSAVSVQVARQ	YEQTVPVPKGS	FYPGETT	106		
QY	116	-----WG----	-----RQ-----	-----QQPPPGGGG	GGGQ-----	-----	133		
DB	107	PLQQLQQGIFWGTSS	QTQGYPSVTS	PRQGSYYFQ	QASFPQPGGQQQ	FGKWKPEGGQQ	166		
QY	134	YYYPYCS--	RPGE	GYGVGGQGR	QRMYP	PCPRGTTGG	PRIGR	172	
DB	167	WYYPSTSLQPGGQ	QITGK--	KQGYPTSLQ	PGQQITGQ	205			

```

RESULT 3
GLT0 WHEAT
ID ID GLT0 WHEAT STANDARD; PRT; 648 AA.
AC AC P10387;
DT 01-WAR-1989 (Rel. 10, Created)
DT 01-WAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutenin, high molecular weight subunit D10 precursor.
GN GLU-D1-2B.
DE OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Poideae;
OC Triticeae; Triticum.
OC NCBI_TaxID=4565;
[1]
RN SEQUENCE FROM N.A.
RP
RP STRAIN=cv. Cheyenne;
RX MEDLINE=89098419; PubMed=2563152;
RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,
RA Malpica-Romero J.M.;
RT "Nucleotide sequences of the two high-molecular-weight glutenin genes
RT from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv
RT Cheyenne."
RL Nucleic Acids Res. 17:461-462(1989).
CC -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -!- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -!- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -!- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPGQG AND
CC GQPGQCGQGYPTS.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X12929; CAA31396.1; --
CC PIR; S04832; S04832.
CC InterPro; IPR003612; AAI.
CC InterPro; IPR001419; Glutenin.
CC Pfam; PF03157; Glutenin_hmw; 1.
CC PRINTS; PR00210; GLUTENIN.
CC SMART; SM00499; AAI; 1.
CC Seed storage protein; Repeat; Multigene family; Signal.
CC SIGNAL
CC 1 21
CC FT CHAIN 22 648 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
CC FT

```

```
FT DOMAIN 147 610 DY10.
SQ SEQUENCE 648 AA; 69629 MW; FE98F1D44B9E9AF1 CRC64;
REPEATS.
Query Match 24.5%; Score 281; DB 1; Length 648;
Best Local Similarity 36.6%; Pred. No. 3.6e-15;
Matches 78; Conservative 23; Mismatches 46; Indels 66; Gaps 11;
QY 1 MAKIAAAAAALCFALVAVAVCOGEVERQRLDLQWQVQSPDLACRQVLDRLQTLGG 60
DB 1 MAKRLVLFVAVV--IALVALTVAEGASRQ----LCERELQESSLEACRQVVDQQL---- 51
QY 61 GGGGGVGPFRWGTLRMRCQQQLQDVSRRCRAIRSMVRYGEE--AMPLEKGMWP----- 115
DB 52 -----AGRLPWSLGLQMRCCQQLRDVSAKRSVAVSQVARYEQTVVPPKGGSYFGETT 106
QY 116 -----WG-----RQ-----QPPPPGGGGGGG----- 133
DB 107 FLQQLQGGIFWGTSSVTGYYPGVTSRQGSYPGQASFPQPGQGGQPKWQEPGQGGQ 166
QY 134 YYYPCS--RPGEYGYGGGGGQRMVPPC--RPG 162
DB 167 WYPTSLQPGGQQQIGK-GQQGYTSLQPG 198
RESULT 4
GLT5 WHEAT
ID GLT5 WHEAT STANDARD; PRT; 839 AA.
AC P10388;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutenin, high molecular weight subunit DX5 precursor.
GN GLU-ID-1D OR GLU-DL-1B.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Cheyenne;
RX MEDLINE=89098419; PubMed=2563152;
RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,
RA Malpica-Romero J.M.;
RT "Nucleotide sequences of the two high-molecular-weight glutenin genes
RT from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv
RT Cheyenne."
RL Nucleic Acids Res. 17:461-462 (1989).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RC STRAIN=cv. Cheyenne;
RA Anderson O.D.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPGQ AND
CC GQPGQGGGYPTS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X12928; CAA31395.1; --
DR
```

```
DR InterPro; IPR001419; Glutenin.
DR Pfam; PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
KW Seed storage protein; Repeat; Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 839 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
FT DOMAIN 131 801 REPEATS.
FT SEQUENCE 839 AA; 89359 MW; OF14E1106D552643 CRC64;
Query Match 13.6%; Score 156.5; DB 1; Length 839;
Best Local Similarity 28.1%; Pred. No. 3.1e-05;
Matches 59; Conservative 20; Mismatches 50; Indels 81; Gaps 11;
QY 1 MAKIAAAAAALCFALVAVAVCOGEVERQRLDLQWQVQSPDLACRQVLDRLQTLGG 57
DB 1 MAKRLVLFVAVV--VALVALTVAEGASEQ----LCERELQELQELKACQVWD---- 51
QY 58 TGGGGGGVGPFRWGTLRMRCQQQLQDVSRRCRAIRSMVRYGEE--AMPLEKGMWP 115
DB 52 -----QQLRDISPECHPVVSPVAGYEQIIVPPKGGSYFP 88
QY 116 -----WG-----RQPPPPGGGGGGGY--YY 136
DB 89 GETTTPQQLQQRIFWGIPIALKRYTSPVTCPOQVSPGQASFPQPGQGGQGGGY 148
QY 137 PCS--RPGEYGYGGGGGQRMVPPC--PCRPG 162
DB 149 PTSPQPGQ--WEEPPGQGGGYTSPQPG 177
RESULT 5
GLT4 WHEAT
ID GLT4 WHEAT STANDARD; PRT; 838 AA.
AC P08489;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutenin, high molecular weight subunit PW212 precursor.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Yamhill;
RX MEDLINE=86093674; PubMed=3001648;
RA Sugiyama T., Rafalski A., Peterson D., Soll D.G.;
RT "A wheat HMW glutenin subunit gene reveals a highly repeated
RT structure."
RL Nucleic Acids Res. 13:8729-8737 (1985).
CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPGQ AND
CC GQPGQGGGYPTS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X03346; CAA27052.1; --
DR PIR; A24107; EEWTHW.
DR InterPro; IPR001419; Glutenin.
```

```
DR Pfam: PF03157; Glutenin_hmw; 1.
DR PRINTS: PR00210; GLUTENIN
KW Seed storage protein; Repeat; Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 838 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
FT DOMAIN 130 799 REPEATS.
FT SEQUENCE 838 AA; 89174 MW; 71D715B7BDF0722D CRC64;
Query Match 13.2%; Score 151.5; DB 1; Length 838;
Best Local Similarity 28.7%; Pred. No. 7.6e-05;
Matches 58; Conservative 18; Mismatches 45; Indels 81; Gaps 10;
QY 1 MAKIAAAAAALFAALVAVAVCOGEVERQLDLOC---WQVQSPDLACHQVLDRLQ 57
DB 1 MAKRLVLFVAV---VALVALTVAGSEASE---LQERELQELQRELKACQVMD--- 51
QY 58 TGGGGGGGGVFRWGTLRMRCQQQLQDVSRCAIRSMVRGYEE----- 104
DB 52 -----QQLRDISPECHPVVSPVAGYEQIIVPKGGSFYPG 88
QY 105 -----AMPPLEKGNP--WGRQQ-----QPPPGGGGGGGYYPYPCSRPG 142
DB 89 ETTTPQLOQRIWFGIPALLKRYVPSVTSPQVSYYPGQASPPRPGQGQ-----QPG 140
QY 143 EGYGYGGGQRMYP--PCRP 162
DB 141 QGQSGGQ-GQGGYPTSPQPG 161
RESULT 6
2SS_BEREX
ID_2SS_BEREX STANDARD; PRT; 146 AA.
AC P04403; P04402;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 2S sulfur-rich seed storage protein precursor (Allergen Ber e 1).
GN BE2S1 AND BE2S2.
OS Bertholletia excelsa (Brazil nut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Lecythidaceae; Bertholletia.
OX NCBI_TaxID=3645;
RN [1]
RP SEQUENCE FROM N.A.
RA Altenbach S.B., Pearson K.W., Leung F.W., Sun S.S.M.;
RT "Cloning and sequence analysis of a cDNA encoding a Brazil nut protein
exceptionally rich in methionine.";
RL Plant Mol. Biol. 8:239-250(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Bassuener R.;
RL Submitted (DEC-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91370890; PubMed=1840683;
RA Gander E.S., Holmstrom K.O., de Paiva G.R., de Castro L.A.B.,
RA Carneiro M., Grossi de Sa M.F.;
RT "Isolation, characterization and expression of a gene coding for a 2S
albumin from Bertholletia excelsa (Brazil nut).";
RL Plant Mol. Biol. 16:437-448(1991).
RN [4]
RP SEQUENCE OF 37-64 AND 70-142.
RX MEDLINE=87004679; PubMed=3759080;
RA Ampe C., van Damme J., de Castro L.A.B., Sampaio M.J.A.M.,
RA van Montagu M., Vandekerckhove J.;
RT "The amino-acid sequence of the 2S sulphur-rich proteins from seeds
of Brazil nut (Bertholletia excelsa H.B.K.).";
RL Eur. J. Biochem. 159:597-604(1986).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE NATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
LINKED BY DISULFIDE BONDS.
```

```
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: M17146; AAA33010.1; -
DR EMBL: X57027; CAA40343.1; -
DR EMBL: X57028; CAA40344.1; -
DR EMBL: X54490; CAA38362.1; -
DR EMBL: X54491; CAA38363.1; ALT_SEQ.
DR EMBL: A13818; CAA01131.1; -
DR PIR: A25802; A25802.
DR PIR: S14946; S14946.
DR PDB: 1GYS; 30-APR-02.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; tryp_alpha_amyl; 1.
DR PRINTS: PR00496; NAPIN.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Signal; Allergen; Pyrrolidone carboxylic acid;
KW 3D-structure.
FT SIGNAL 1 22
FT PROPEP 23 36 SMALL CHAIN.
FT CHAIN 37 64
FT PROPEP 65 69 LARGE CHAIN 1B.
FT CHAIN 70 142
FT PROPEP 143 146
FT MOD_RES 37 37
FT VARIANT 91 91
FT CONFLICT 38 39 S -> E (IN VARIANT 1A).
FT CONFLICT 122 122 E -> Q (IN REF. 4).
FT CONFLICT 126 126 L -> M (IN REF. 4).
FT CONFLICT 126 126 I -> L (IN REF. 4).
SQ SEQUENCE 146 AA; 16911 MW; A7DF778FD766410D CRC64;
Query Match 10.0%; Score 114.5; DB 1; Length 146;
Best Local Similarity 28.7%; Pred. No. 0.012;
Matches 37; Conservative 20; Mismatches 33; Indels 39; Gaps 7;
QY 1 MAKIAAAAAA-----ALCFALVAVAV-----COGEVERQLRD---LQWQE 40
DB 1 MAKISVAAAALLVLMALGHATAFRATVTTTVEENQEECEQMQRQQLSHCRMTMRQQ 60
QY 41 VQESPLDACRQVLDRLQTLGGGGGGVGPFRWGTLRMRCQQQLQDVSRCAAIR-SMV 99
DB 61 MEESPY-----QTMPPR-----MSECEQLEGWDESCRCGLRMMMM 102
QY 100 RGYEEAMPP 108
DB 103 RMQSEMQP 111
RESULT 7
GDA4 WHEAT
ID GDA4 WHEAT STANDARD; PRT; 297 AA.
AC P04724;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Alpha/beta-gliadin A-IV precursor (Prolamin).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85234522; PubMed=2989281;
RA Okita T.W., Cheesbrough V., Reeves C.D.;
```


"Evolution and heterogeneity of the alpha-/beta-type and gamma-type gliadin DNA sequences.";
J. Biol. Chem. 260:8203-8213(1985).
-|- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
-|- MISCELLANEOUS: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5 HOMOLOGOUS CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M11075; AAA34282.1; -;
PIR; T06500;
InterPro; IPR003612; AAI.
Pfam; PF00234; tryp_alpha_aml1; 1.
DR SMART; SM00499; AAI; 1.
KW Seed storage protein; Repeat; Signal; Multigene family.
FT SIGNAL 1 20
FT CHAIN 21 297 ALPHA/BETA-GLIADIN A-IV
FT SEQUENCE 297 AA; 34239 MW; 002SED89AE9588B CRC64;

Query Match 9.9%; Score 114; DB 1; Length 297;
Best Local Similarity 23.0%; Pred. No. 0.026;
Matches 44; Conservative 29; Mismatches 76; Indels 42; Gaps 6;

QY 24 QGEVQRRLDLQWQEVQSPDLACRQVLDRLQTLGGGGGGVGFPRMGTLRMRCCQQL 83
DB 125 QQQ 184
QY 84 QDVSRRCRCRAIRSMVRG-----YEEAMPPEKGMWPMGRQOQPPPGGGGG 130
DB 185 WQIPEGRQCAIRNVHAIILHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 238
QY 131 QGGYYPPSPRGYGGYGGGQ-GRQMYPPCRPGTGGGPRGRVRLTKAREYA-AGLPMW 188
DB 239 QGSF-----QPSQNPQAGSQVQPQLP-----QPEIRNLAEITLPM 277
QY 189 CRLSEFOECI 199
DB 278 CNYVIPPYCTI 288

RESULT 8
ID 2SS_RICCO STANDARD; PRT; 258 AA.
AC P01089;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2S albumin precursor (Allergen Ric c 1).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91109729; PubMed=2274038;
RA Irwin S.D., Keen J.N., Findlay J.B.C., Lord J.M.;
RT "The Ricinus communis 2S albumin precursor: a single preproprotein may be processed into two different heterodimeric storage proteins.";
RL Mol. Gen. Genet. 222:400-408(1990).
RN [2]

RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91016940; PubMed=2216785;
RA Irwin S.D., Lord J.M.;
RT "Nucleotide sequence of a Ricinus communis 2S albumin precursor gene.";
RL Nucleic Acids Res. 18:5890-5890(1990).
RN [3]
RP SEQUENCE OF 157-190 AND 194-258.
RX MEDLINE=83082772; PubMed=7174664;
RA Sharief F.S., Li S.S.-L.;
RT "Amino acid sequence of small and large subunits of seed storage protein from Ricinus communis.";
RL J. Biol. Chem. 257:14753-14759(1982).
RN [4]
RP SIMILARITY TO PROTEINASE INHIBITORS.
RX MEDLINE=83308577; PubMed=6615448;
RA Odani S., Koide T., Ono T., Ohnishi K.;
RT "Structural relationship between barley (Hordeum vulgare) trypsin inhibitor and castor-bean (Ricinus communis) storage protein.";
RL Biochem. J. 213:543-545(1983).
CC -|- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -|- SUBUNIT: THE NATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN LINKED BY 2 DISULFIDE BONDS.
CC -|- PTM: THE TWO DISULFIDE BONDS LINKING THE SMALL AND THE LARGE CHAINS INVOLVE CYS-162 AND CYS-175.
CC -|- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X54158; CAA38097.1; -;
PIR; S11499; RZCS.
InterPro; IPR003612; AAI.
InterPro; IPR000617; Napin.
Pfam; PF00234; tryp_alpha_aml1; 2.
DR PRINTS; PR00496; NAPIN.
DR SMART; SM00499; AAI; 1.
KW Seed storage protein; Signal; Allergen; Pyrrolidone carboxylic acid.
FT SIGNAL 1 21
FT PROPEP 22 156
FT CHAIN 157 190 2S ALBUMIN, SMALL CHAIN.
FT CHAIN 191 193
FT CHAIN 194 258 2S ALBUMIN, LARGE CHAIN.
FT MOD RES 194 194 PYRROLIDONE CARBOXYLIC ACID.
FT CONFLICT 222 222 E -> Q (IN REF. 3).
FT CONFLICT 226 229 MISSING (IN REF. 3).
FT CONFLICT 234 234 D -> N (IN REF. 3).
FT CONFLICT 255 255 E -> Q (IN REF. 3).
SQ SEQUENCE 258 AA; 29290 MW; 27874CFC50E41072 CRC64;

Query Match 9.8%; Score 112.5; DB 1; Length 258;
Best Local Similarity 21.7%; Pred. No. 0.03;
Matches 59; Conservative 34; Mismatches 88; Indels 91; Gaps 11;

QY 1 MAKI-AAAAAALCF-----AALVAVACGVEVERLRDLQCMQVQESPLDA 48
DB 1 MAKLPITIALVSVLLFIANASFAVTTTITTEIDSKGEREGSSSQCRQEVQRDLSS 60
QY 49 CRQVLDRLQTLGGGGGGVGFPRMGTLRM-----RCCQQLQDVSRRCRAA 94
DB 61 CERYLRFQSSRRSPGSEV-----LRMPDGNQOQESQQLQCCCNQVQRDEQCEA 112
QY 95 ISMSRWGYEAMPPEKGMWPMGRQOQPPPGGGGGGGYVPCSR-----PGEYGYGQ 149
DB 113 IKYI-----AEDQIQQQL-HGESERVAORAGEIVSSCGVRCMRQTRTNPSQOGRGQ 165
QY 150 GGORQWYPPC---RGTGGGPR----- 169

```
Db 166 IQSQNLRLQCOEYIKQVSGQRRSDNQBSLGGCDHLKQMSQRCBGLRQALEQQQ 225
QY 170 -IGRVR---ITKAREYAAGLPMCRSLSEPOEC 197
Db 226 SQGLOQDVFEAFRTAANLPSMCGVS-PTEC 256

RESULT 9
2SS5_HELAN STANDARD; PRT; 295 AA.
AC P15461;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2S seed storage protein precursor (2S albumin storage protein).
GN HAG5.
OS Helianthus annuus (Common sunflower).
OC Spermatophyta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Asteridae; Campanulids; Asterales; Asteraceae; Core eudicots;
OC Heliantheae; Helianthus.
OX NCBI_TaxID=4232;
RN 1;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 162-173.
RC STRAIN=cv. Giant grey stripe.
RX MEDLINE=88142538; PubMed=2830455;
RA Allen R.D., Cohen E.A., Vonder Haar R.A., Adams C.A., Ma D.P.,
RT Nessler C.L., Thomas T.L.;
RA "Sequence and expression of a gene encoding an albumin storage
  protein in sunflower.";
RL Mol. Gen. Genet. 210:211-218 (1987).
CC -!- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -!- PTM: IT IS POSSIBLE THAT THE 38 KDA PRECURSOR IS CLEAVED INTO TWO
  POLYPEPTIDES THAT ARE APPROXIMATELY THE SAME SIZE. THE MATURE
  PROTEIN IS COMPOSED OF A SINGLE POLYPEPTIDE CONTAINING ONE OR MORE
  INTRA-MOLECULAR DISULFIDE LINKAGES.
CC -!- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X06410; CAA29699.1; -
DR PIR; S01062; S01062.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; tryp_alpha_amyl; 2.
DR SMART; SM00499; AAI; 2.
KW Seed storage protein; Signal; Multigene family.
FT SIGNAL 1 20
FT PROPEP 21 163
FT CHAIN 162 295
SQ SEQUENCE 295 AA; 34071 MW; 8958A106805142A1 CRC64;

Query Match 9.8%; Score 112; DB 1; Length 295;
Best Local Similarity 22.0%; Pred. No. 0.037;
Matches 38; Conservative 21; Mismatches 54; Indels 60; Gaps 4;

QY 36 QCHQVEQESPLDACRVLDRQLTGCGGGGCVGPRWG-----TGLMRCCQQLQDVSRRE 89
Db 177 QCRETEIQRVSGRCORFVEQMOQSPRSTRPYOQRPGQQQQQQOQRLQCCCNLQNVKRE 236
QY 90 CRCAAIRSGVGEAEAMPLEKGMWPGWGRQQPPQGGGGGQGYYPYPCSRPGEVGYGQ 149
Db 237 CHCEAIQEVARRVNRQ-----PQQQQQRRGQFGGQ----- 267
QY 150 GGQRQMYPPCRPGTGGPRIGRVLTKAREYAAGLPMCRSLSEPOECIFSG 202
Db 268 -----EMETARRVIQLNLPNQCDL-EVQQCTTCG 295

RESULT 10
GDA9_WHEAT STANDARD; PRT; 307 AA.
AC P18573;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Alpha/beta-gliadin MM1 precursor (prolamin).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Chinese Spring; TISSUE=Endosperm;
RX MEDLINE=91346679; PubMed=2102865;
RA Garcia-Maroto F., Manana C., Garcia-Olmedo F., Carbonero P.;
RT "Nucleotide sequence of a cDNA encoding an alpha/beta-type gliadin
  from hexaploid wheat (Triticum aestivum).";
RL Plant Mol. Biol. 14:867-868 (1990).
CC -!- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC -!- MISCELLANEOUS: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5
  HOMOLOGY CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE
  TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS
  WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN 100 COPIES OF
  THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).

EMBL; X17361; CAA35238.1; -
DR PIR; S10015; S10015.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001954; Gliadin.
DR Pfam; PF00234; tryp_alpha_amyl; 1.
DR PRINTS; PR00208; GLIADGLUTEN.
DR SMART; SM00499; AAI; 1.
KW Seed storage protein; Repeat; Signal; Multigene family.
FT SIGNAL 1 20
FT CHAIN 21 307
SQ SEQUENCE 307 AA; 35397 MW; 06C1858BD96F1E08 CRC64;

Query Match 9.7%; Score 111.5; DB 1; Length 307;
Best Local Similarity 23.4%; Pred. No. 0.043;
Matches 44; Conservative 29; Mismatches 76; Indels 39; Gaps 6;

QY 24 QGVEVERORLRLQCHQVEQESPLDACRVLDRQLTGCGGGGCVGPRWGTLGRMRCCQQL 83
Db 138 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 197
QY 84 QDVSRRCRCARSRMVRG-----YEEAMPLEKGMWPGWGRQQPPQGGGGGQGG 133
Db 198 WQIPQSRQCAIHNVVHAILLHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 251
QY 134 YYPYPCSRPGEVGYGCGG--QRQMYPPCRPGTGGPRIGRVLTKAREYA--AGLPMCR 191
Db 252 F-----QFSQQNPQAGSVQPPQLP-----QFEIRNLAEITLPMCNV 290

192 SEPQECST 199
291 YIPPYCTI 298

RESULT 11
GDA6_WHEAT
```

```

ID  GDAG6 WHEAT STANDARD; PRT; 296 AA.
AC  F04726;
DT  13-AUG-1987 (Rel. 05, Created)
DT  13-AUG-1987 (Rel. 05, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  Alpha/beta-gliadin clone PW1215 precursor (Prolamin).
OS  Triticum aestivum (Wheat).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC  Triticeae; Triticum.
OX  NCBI_TaxID=4565;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=85242077; PubMed=38939304;
RA  Summer-Smith M., Rafalski J.A., Sugiyama T., Stoll M., Soell D.;
RT  "Conservation and variability of wheat alpha/beta-gliadin genes.";
RL  Nucleic Acids Res. 13:3905-3916(1985).
CC  -!- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC  -!- MISCELLANEOUS: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5
CC  HOMOLOGUE CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE
CC  TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS
CC  WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF
CC  THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X02538; AAA26383.1; -.
DR  EMBL; K01074; AAA34277.1; -.
DR  PIR; S07361; S07361.
DR  InterPro; IPR003612; AAI.
DR  Pfam; PF00234; tryp_alpha_aml; 1.
DR  PRINTS; PR00208; GLIADGLUTEN.
DR  SMART; SM00499; AAI; 1.
KW  Seed storage protein; Repeat; Signal; Multigene family.
FT  SIGNAL 1 20
FT  CHAIN 21 296 ALPHA/BETA-GLIADIN CLONE PW1215.
SQ  SEQUENCE 296 AA; 33941 MW; A9BDF590AD40F135 CRC64;

Query Match 9.7%; Score 111; DB 1; Length 296;
Best Local Similarity 23.2%; Pred. No. 0.045;
Matches 43; Conservative 26; Mismatches 76; Indels 40; Gaps 6;

QY  24 QGEVERQLRLDLCMQEVOESPLDACRQVLDRLQTLGGGGGGVGPFRWGTLRMRCQQQL 83
Db  130 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 189
QY  84 QDVSRRCRAAIRSMVRG---YEEAMPPLEKGMWPGWROQQPPPPGGGGGGGGYPCSR 140
Db  190 WQIPESRCQAHHNVHAILHQQQ-----RQQPSSQ-----VSLQQ 227
QY  141 PGEICYGCGQGGQGMYPGPRGTTGG-----PRIGRVLTKAREYAAGLPMWRLSE 193
Db  228 PQQQVPSGGG-----FPQSQNPQAGSQVQPOLPQFEIRNLALQT-----LPRMCNVI 279
QY  194 PQECS 198
Db  280 PYCS 284

RESULT 12
GLT1 WHEAT
ID  GLT1 WHEAT STANDARD; PRT; 101 AA.
AC  F02861;
DT  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DT  01-OCT-1994 (Rel. 30, Last annotation update)

```

```

DE  Glutenin, high molecular weight subunit PC256 (Fragment).
OS  Triticum aestivum (Wheat).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC  Triticeae; Triticum.
OX  NCBI_TaxID=4565;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Forde J., Forde B.G., Fry R.P., Kreis M., Shewry P.R., Miflin B.J.;
RT  "Identification of barley and wheat cDNA clones related to the high-
RT  M-r polypeptides of wheat gluten.";
RL  FEBS Lett. 162:360-366(1983).
CC  -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC  PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC  VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC  -!- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC  -!- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC  GROUP 1 CHROMOSOMES OF WHEAT.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X00054; CAA24933.1; -.
DR  PIR; A03352; SEWTL.
DR  InterPro; IPR001419; Glutenin.
DR  Pfam; PF03157; Glutenin hmw; 1.
DR  PRINTS; PR00210; GLUTENIN.
KW  Seed storage protein; Repeat; Multigene family.
FT  NON_TER 1
SQ  SEQUENCE 101 AA; 10896 MW; EFD6ADDDPEED993 CRC64;

Query Match 9.6%; Score 110; DB 1; Length 101;
Best Local Similarity 33.3%; Pred. No. 0.02;
Matches 36; Conservative 11; Mismatches 25; Indels 36; Gaps 6;

QY  117 GRQQQP-----PPQGGGGGGGGYVYPCSS--RFGEGYGVGGGQGMYPGPRGTTGGP- 168
Db  4 GCGQGPQWLQPRQCGQ-----YYTSPQSGCGQQLGQG--QQGYTSPQSGCGQ 56
QY  169 -----RIGRVLTKAREYAAGLPMWRLSEPCSECSIFSGD 204
Db  57 GYDSPYHVSAEHQAAASLKVAKAQAQLPAMCRL-----EGGD 95

RESULT 13
IA02 WHEAT
ID  IA02 WHEAT STANDARD; PRT; 145 AA.
AC  P16851;
DT  01-AUG-1990 (Rel. 15, Created)
DT  01-MAR-1992 (Rel. 21, Last sequence update)
DT  01-MAR-1992 (Rel. 21, Last annotation update)
DE  Alpha-amylase/trypsin inhibitor CM2 precursor (Chloroform/methanol-
DE  soluble protein CM2).
OS  Triticum aestivum (Wheat).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC  Triticeae; Triticum.
OX  NCBI_TaxID=4565;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=cv. Agathe; TISSUE=Seed;
RX  MEDLINE=91370875; PubMed=1893104;
RA  Gautier M.F., Alary R., Lullien V., Joudrier P.;
RT  "Nucleotide sequence of a cDNA clone encoding the wheat (Triticum
RT  durum Desf.) CM2 protein.";
RL  Plant Mol. Biol. 16:333-334(1991).
RN  [2]
RP  SEQUENCE OF 26-53.

```


Db 264 GWNQ-----GSGGGP 275

RESULT 15
GDA2 WHEAT

ID_GDA2_WHEAT STANDARD; PRT; 291 AA.

AC P04722;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Alpha/beta-gliadin A-II precursor (Prolamin).
OS Triticum aestivum Wheat.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85234522; PubMed=2989281;
RA Okita T.W., Cheesbrough V., Reeves C.D.;
RT "Evolution and heterogeneity of the alpha-/beta-type and gamma-type
gliadin DNA sequences";
RL J. Biol. Chem. 260:8203-8213 (1985).
CC -!- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC -!- MISCELLANEOUS: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5
CC HOMOLOGY CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE
CC TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS
CC WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF
CC THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M10092; AAA34276.1; -;
DR PIR; T06498; T06498.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001954; GliA glutenin.
DR Pfam; PF00234; tryf_alpha_aml; 1.
DR PRINTS; PR00208; GLIADGLUTEN.
DR SMART; SM00499; AAI; 1.
KW Seed storage protein; Repeat; Signal; Multigene family.
FT SIGNAL 1 20
FT CHAIN 21 291 ALPHA/BETA-GLIADIN A-II.
SQ SEQUENCE 291 AA; 33661 MW; 9839F93B0825A2E0 CRC64;

Query Match 9.5%; Score 109.5; DB 1; Length 291;
Best Local Similarity 21.6%; Pred. No. 0.058;
Matches 41; Conservative 30; Mismatches 80; Indels 39; Gaps 4;

QY 24 QGEVERQLRLDQCWQEVQSPLDACRQVLDRLQITGGGGGGVGFPRWGTGLMRCCQQL 83
Db 118 QQQQQQQQQQQQILQILQQLIPCRDVLQHNTHAGSSQVLQSTYQLVQLCCQQL 177
QY 84 QDVRECRCAAIRSMVRG-----YEEAMPLEKGMWPMGRQQQPPQGGG 129
Db 178 WQIPEQSRCAIHNVVHAIILHQHHHHQQQQQQQQQPLSQVFSQQPQQYPSGQ---- 233
QY 130 GQGGYYPCSRPEGYGQGGQRYPPCRPGTGGGPRIGRVLTAKREYAAGLPMWC 189
Db 234 ---GFFQPSQNNFQAQGSFQQL-----PQFEIRNALQT----LPAMC 272
QY 190 RLSEPOECISI 199
Db 273 NVVIPPVCTI 282

Search completed: November 29, 2003, 17:05:29
Job time : 23 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2003, 16:53:04 ; Search time 56 Seconds
(without alignments)
949.265 Million cell updates/sec

Title: US-10-053-410-4
Perfect score: 1148
Sequence: 1 MAKIAAAAAALCFALVAV.....MMCRSLSEPOECIFSGDQY 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_protein:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1148	100.0	206	10	Q946V3 zea mays (m
2	439	38.2	186	10	P93414 oryza sativ
3	414	36.1	226	10	Q8LKV8 aegilops ta
4	292.5	25.5	654	10	Q93XR5 aegilops cy
5	289	25.2	720	10	Q941J6 triticum ae
6	288.5	25.1	624	10	Q8LKV7 aegilops ta
7	287	25.0	705	10	Q03871 triticum ae
8	283	24.7	704	10	Q8LKI8 aegilops sp
9	282.5	24.6	475	10	Q40045 hordeum vul
10	282	24.6	648	10	Q38767 aegilops ta
11	281.5	24.5	707	10	Q40054 aegilops ta
12	277.5	24.2	196	10	Q8H0L6 triticum ae
13	276.5	24.1	655	10	Q8S3V7 aegilops um
14	275	24.0	713	10	Q94IK7 secale cere
15	267.5	23.3	737	10	Q94IL4 secale cere
16	267	23.3	666	10	Q93XR6 aegilops cy

17	267	23.3	707	10	Q94IL6 secale cere
18	267	23.3	713	10	Q94IJ8 triticum ae
19	267	23.3	713	10	Q9SDM2 triticum ae
20	267	23.3	713	10	Q93WM1 secale cere
21	264.5	23.0	179	10	Q9M5N3 aegilops ve
22	260	22.6	633	10	Q8S3V9 aegilops ma
23	259	22.6	713	10	Q94IK8 secale cere
24	257	22.4	181	10	Q8H0L3 triticum ae
25	257	22.4	587	10	Q93XJ7 triticum ti
26	253	22.0	713	10	Q94IL1 secale cere
27	172.5	15.0	743	10	Q94IL0 secale cere
28	166	14.5	754	10	Q93WFO secale cere
29	166	14.5	754	10	Q94IK9 secale cere
30	165	14.4	754	10	Q94IL2 secale cere
31	165	14.4	766	10	Q9SDM3 triticum ae
32	164	14.3	462	10	Q94IJ9 triticum ae
33	163.5	14.2	818	10	Q93XJ8 triticum ti
34	162	14.1	971	10	Q8GV12 triticum ae
35	160.5	14.0	238	10	Q8H0L2 triticum ae
36	160.5	14.0	811	10	Q94IL5 secale cere
37	160.5	14.0	811	10	Q94IJ7 triticum ae
38	158.5	13.8	836	10	Q8S3T2 aegilops ta
39	158.5	13.8	836	10	Q8S3T3 aegilops ta
40	158	13.8	845	10	Q8LKV6 aegilops ta
41	157	13.7	204	10	Q9M5N2 aegilops ve
42	157	13.7	405	10	Q41516 triticum ae
43	157	13.7	405	10	Q9SYV0 triticum ae
44	157	13.7	766	10	Q94IL3 secale cere
45	157	13.7	830	10	Q03872 triticum ae

ALIGNMENTS

RESULT 1

Q946V3 PRELIMINARY; PRT; 206 AA.

AC Q946V3; TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Alpha globulin.

OS Zea mays (Maize).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC PACCAD clade; Panicoideae; Andropogoneae; Zea.

OX NCBI_TaxID=4577;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Endosperm;

RX MEDLINE=21480069; PubMed=11595803;

RA Woo Y.M., Hu D.W., Larkins B.A., Jung R.;

RT "Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression."

RL Plant Cell 13:2297-2317(2001).

DR EMBL; AF371278; RAL16993.1; -

DR InterPro; IPR003612; AAI.

DR Pfam; PF00234; tryp_alpha_amyl; 1.

DR SMART; SM00499; AAI; 1.

SQ SEQUENCE 206 AA; 22298 MW; B146CCA41BD82089 CRC64;

Query Match 100.0%; Score 1148; DB 10; Length 206;
Best Local Similarity 100.0%; Pred. No. 1.5e-105;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAKIAAAAAALCFALVAVVAVCQGEVERQRDLQCWQEVQESPLDACRQLDRLTGG	60
DB	1	MAKIAAAAAALCFALVAVVAVCQGEVERQRDLQCWQEVQESPLDACRQLDRLTGG	60
QY	61	GGGGVGPPFRWGTGLRMRCQQQLQDVSRRCRCACAIIRSMVRGYEAMPPLKGMWPGRQ	120
DB	61	GGGGVGPPFRWGTGLRMRCQQQLQDVSRRCRCACAIIRSMVRGYEAMPPLKGMWPGRQ	120

```

QY 121 QPPQGGGGGGGGYPCSRPGEYGGGGGQGMYPPCRPGTTGGPRIGRVRLLTKARE 180
DB |||||
QY 121 QPPQGGGGGGGGYPCSRPGEYGGGGGQGMYPPCRPGTTGGPRIGRVRLLTKARE 180
DB |||||

181 YAAGLPMWRLSPQPCSFIFSGDDQY 206
181 YAAGLPMWRLSPQPCSFIFSGDDQY 206
DB |||||

RESULT 2
P93414 PRELIMINARY; PRT; 186 AA.
ID P93414;
AC P93414;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 26 kDa globulin.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=96235139; PubMed=8666249;
RX Nakase M., Hotta H., Adachi T., Aoki N., Nakamura R., Masumura T.,
RA Tanaka K., Motoda T.;
RT "Cloning of the rice seed alpha-globulin-encoding gene: sequence
RT similarity of the 5'-flanking region to those of the genes encoding
RT wheat high molecular-weight glutenin and barley D hordein.";
RL Gene 170:223-226(1996).
DR EMBL; D50643; BAA09308.1; -.
DR Gramene; P93414; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001419; Glutenin.
DR Pfam; PF00234; tryd alpha amyl; 1.
DR PRINTS; PR00210; GLUTENIN.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 186 AA; 21055 MW; AE2B8F1107C8BC94 CRC64;

Query Match 38.2%; Score 439; DB 10; Length 186;
Best Local Similarity 51.4%; Pred. No. 1.5e-35;
Matches 108; Conservative 20; Mismatches 54; Indels 28; Gaps 9;

QY 1 MAKIAAAAAALCFALVAVAVCGEVEERQLDLQCWQVQESPLDACKQVLDRLQTGG 60
DB |||||
QY 1 MASKVWFFFAAL-MAAMVAISGAQLSESEMRFRDQCQREVQDSPLDACKQVLDRLQTR 59
DB |||||

61 GGGGGVGFPRWGTGLMRCCQQLQDVSRCCAAIRSMVRGVEEAMP-PLEKGNWPGWGRQ 119
DB |||||
60 ERFPQMFRRFGALGLRMCCQQLQDVSRCCAAIRSMVRGVEEAMP-PLEKGNWPGWGRQ 119
DB |||||

120 QPPQGGGGGGGGYPCSRPGEYGGGGGQGMYPPCRPGTTGGPRIGRVRLLT 176
DB |||||
120 YY-----GGEG-----SSSQGY-YGEGSSEGYGQQQPGMT-----RVRLT 158
DB |||||

177 KAREYAAGLPMWRLSPQPCSFIFSGDDQY 206
DB |||||
159 RARQYAAQLPSMCKV-EPQCCSIFAAG-QY 186
DB |||||

RESULT 3
Q8LKV8 PRELIMINARY; PRT; 226 AA.
ID Q8LKV8;
AC Q8LKV8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Seed globulin.
OS Aegilops tauschii (Aegilops squarrosa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Aegilops.
OX NCBI_TaxID=37882;
RN [1]
RP SEQUENCE FROM N.A.
RA Anderson O.D., Rausch C., Mouillet O., Lagudah E.S.;
RT "Characterization of a wheat D-genome BAC containing two paralogous
RT HMW-glutenin genes: distribution of genes and retrotransposon
RT clusters.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF497474; AAM77580.1; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001419; Glutenin.
DR Pfam; PF00234; tryd alpha amyl; 1.
DR PRINTS; PR00210; GLUTENIN.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 226 AA; 24515 MW; 95397E83C63D50AD CRC64;

Query Match 36.1%; Score 414; DB 10; Length 226;
Best Local Similarity 46.8%; Pred. No. 5.5e-33;
Matches 110; Conservative 21; Mismatches 48; Indels 56; Gaps 14;

QY 11 ALCFAALVAVAVCGEVEERQLDLQCWQVQESPLDACKQVLDRLQTGGGGGGYV-PF 69
DB |||||
8 AVFLAALVAGSAAQGVLE-QSLTDAQCRGEVQEKPLACRQLLEQQLTGRAGEGAVGVPL 66
DB |||||

70 ---RWGTGLMRCCQQLQDVSRCCAAIRSMVRGVEEAMPLEKG-WHPWGRQ--- 121
DB |||||
67 FPAQWGA--RERCCRLQESVSRCCAAIRSMVRDYEQSMPLGEGRHSGSGERQEQGC 124
DB |||||
122 -----PPPO-----GGGGGGGGYPCSRPGEYGGGGGQGMYPPCR--RP 161
DB |||||
125 SGESTEPEQRQEVQGGYSGTGGGQGGGTH-----GTVRGGGQRQGVLCHEKP 177
DB |||||

162 -----GTTGGG-----PRIGRVRLLTKAREYAAGLPMWRLSPQPCSFIFSGDDQY 206
DB |||||
178 QRQCGEGFSGEAGAQKQKQAGVRLLTKVR-----LPTACRI-EPQECVFT-ADQY 225
DB |||||

RESULT 4
Q93XR5 PRELIMINARY; PRT; 654 AA.
ID Q93XR5;
AC Q93XR5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glutenin HMW subunit 1Dy precursor (Fragment).
GN GLU-1D-2.
OS Aegilops cylindrica.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Aegilops.
OX NCBI_TaxID=130456;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RA Wan Y., Wang D., Shewry P.R., Halford N.G.;
RT "Isolation and characterization of five novel high molecular weight
RT subunit of glutenin genes from Triticum timopheevi and Aegilops
RT cylindrica.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ306974; CAC38033.1; -.
DR InterPro; IPR001419; Glutenin.
DR Pfam; PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
KW Signal.
FT SIGNAL 1 21
FT CHAIN 22 >654 HMW SUBUNIT 1DY.
FT NON_TER 654 654
SQ SEQUENCE 654 AA; 70473 MW; AC8E71075FCFD2C7 CRC64;

Query Match 25.5%; Score 292.5; DB 10; Length 654;
Best Local Similarity 36.2%; Pred. No. 1.8e-20;

```



```
Matches 80; Conservative 25; Mismatches 51; Indels 65; Gaps 10;
QY 1 MAKIAAAAAALCFALVAVVCGQVEVERQLRDLCWOEVSPLDACRQVLDRLTGG 60
Db 1 MAKRLVLFRAVV--IALVALTTAEASRQ-----LQERELQESSLEACRQVVDQQL--- 51
QY 61 GGGGGVGPFRWGTGLRMCCQQLQDVSRRCRAIRSMVRYEE--AMPPLEKGMWP----- 115
Db 52 -----AGRLPWSGTGLMRCQQLRDVSAKRSVAVSQVTRQYEQTVVPPKGRSFYPGETT 106
QY 116 -----WG-----RQ-----QQPPPGGGGGGG----- 133
Db 107 PLQQLQQGIFWGTSSQTQVGYPSVTSRQGSYPGQASFPQPGQPGKMQEPGQGO 166
QY 134 YYPYPC--RPGEYGYGQGGQQRQMPYPCRPPTGGGPRGR 172
Db 167 WYPTSLQPGGQGIKGG--QQGYPTSLQPGGQGOIIGQ 205

RESULT 5
Q941J6 PRELIMINARY; PRT; 720 AA.
AC Q941J6;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DE 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE High molecular weight glutenin subunit y precursor.
GN GLU-IR.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RL encoded by Glu-1R gene of Secale cereale.";
RN Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314785; CAC40687.1; -.
DR InterPro; IPR001419; Glutenin.
DR Pfam; PF03157; Glutenin hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
KW SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT Y.
FT SIGNAL 22 720
SQ SEQUENCE 720 AA; 77372 MW; 3385D2004B7D3B18 CRC64;

Query Match 25.2%; Score 289; DB 10; Length 720;
Best Local Similarity 32.9%; Pred. No. 4.4e-20;
Matches 85; Conservative 24; Mismatches 47; Indels 102; Gaps 13;
QY 17 LVAVVCGQVEVERQLRDLCWOEVSPLDACRQVLDRLTGGGGGGVGPFRWGTGLR 76
Db 15 LVALTAEGEASRQ-----LQERELQESSLEACRQVVDQQL-----AGRLPWSGTGLQ 62
QY 77 MRCCQQLQDVSRRCRAIRSMVRYEE--AMPPLEKGMWP-----WGR--- 118
Db 63 MRCCQQLRDVSAKRLVAVSQVTRQYEQTVVPPKGRSFYPGETTPLQQLQVIFWGTSSQ 122
QY 119 -----QQP-----PPGG-----GGGGGGY----- 134
Db 123 TVGYPSVSSPQGYYPGQASFPQPGQPGKQWELGGQGGYPTSLHSGGGQGG 182
QY 135 YYPYPC--RPGEYGYGQGG-----GQRMYP--PCRPG---TTGGG 167
Db 183 YPSSSLQPGGQGIKGGQGGYPTSLQPGGQGOIIGQGGYPTSPQHPGQRQPGQG 242
```

```
QY 168 PRGR-VRLTKAREYAAAG 184
Db 243 QQIGGQQQLGGQRIGQG 260

RESULT 6
Q8LKV7 PRELIMINARY; PRT; 624 AA.
AC Q8LKV7;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE HMW-glutenin.
GN GLU-DT.
OS Aegilops tauschii (Aegilops squarrosa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Aegilops.
OX NCBI_TaxID=37682;
RN [1]
RP SEQUENCE FROM N.A.
RA Anderson O.D., Rausch C., Mouillet O., Lagudah E.S.;
RT "Characterization of a wheat D-genome BAC containing two paralogous
RT HMW-glutenin genes: distribution of genes and retrotransposon
RT clusters.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF497474; AAM77581.1; -.
DR InterPro; IPR001419; Glutenin.
DR Pfam; PF03157; Glutenin hmw; 3.
DR PRINTS; PR00210; GLUTENIN.
SQ SEQUENCE 624 AA; 67104 MW; 9EA4B0C0B41CF829 CRC64;

Query Match 25.1%; Score 288.5; DB 10; Length 624;
Best Local Similarity 36.2%; Pred. No. 4.2e-20;
Matches 80; Conservative 24; Mismatches 52; Indels 65; Gaps 10;
QY 1 MAKIAAAAAALCFALVAVVCGQVEVERQLRDLCWOEVSPLDACRQVLDRLTGG 60
Db 1 MAKRLVLFRAVV--IALVALTTAEASRQ-----LQERELQESSLEACRQVVDQQL--- 51
QY 61 GGGGGVGPFRWGTGLRMCCQQLQDVSRRCRAIRSMVRYEE--AMPPLEKGMWP----- 115
Db 52 -----AGRLPWSGTGLMRCQQLRDVSAKRSVAVSQVTRQYEQTVVPPKGRSFYPGETT 106
QY 116 -----WG-----RQ-----QQPPPGGGGGGG----- 133
Db 107 PLQQLQQGIFWGTSSQTQVGYPSVTSRQGSYPGQASFPQPGQPGKMQEPGQGO 166
QY 134 YYPYPC--RPGEYGYGQGGQQRQMPYPCRPPTGGGPRGR 172
Db 167 WYPTSLQPGGQGIKGG--QQGYPTSLQPGGQGOIIGQ 205

RESULT 7
Q03871 PRELIMINARY; PRT; 705 AA.
AC Q03871;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE HMW glutenin subunit 1By9 precursor.
GN GLU-1BY9.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Cheyenne; TISSUE=Endosperm;
RA Halford N.G.;
RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
```

```

RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=cv. Cheyenne; TISSUE=Endosperm;
RA  Halford N.G.; Forde J.; Anderson O.D.; Greene F.C.; Shewry P.R.;
RT  "The nucleotide and deduced amino acid sequences of an HMW glutenin
RT  subunit gene from chromosome 1B of bread wheat (Triticum aestivum L.)
RT  and comparison with those of genes from chromosomes 1A and 1D.";
RL  Theor. Appl. Genet. 75:117-126(1987).
DR  EMBL; X61026; CAA43361.1; -.
DR  InterPro; IPR003612; AAI.
DR  Pfam; PF03157; Glutenin_hmw; 1.
DR  PRINTS; PR00210; GLUTENIN.
DR  SMART; SM00499; AAI; 1.
KW  Signal.
FT  SIGNAL.
FT  CHAIN.
SQ  SEQUENCE 705 AA; 75702 MW; 88BC745940710FF8 CRC64;

Query Match 25.0%; Score 287; DB 10; Length 705;
Best Local Similarity 32.9%; Pred. No. 6.8e-20;
Matches 85; Conservative 24; Mismatches 47; Indels 102; Gaps 13;

QY 17 LVAVAVCOGVEVERQLRDLCQWQVQESPLDACRQVLDRLTGGGGGGVGPFRWGTGLR.76
DB 15 LVALTAEGEASRQ-----LQCERELQESSLEACRQVVDQQL-----AGRLPWSGTGLQ 62

QY 77 MRCCQQLQDVSRRCRAAIRSMVRGYE-AMPPLEKGMWP-----WGR--- 118
DB 63 MRCCQQLRDVSAKCRPVAIVQARQYEQTVVPKGGSFYPGETTPLOQLQGVIFWGTSSQ 122

QY 119 -----QQQP-----PPQG-----GGGGGY----- 134
DB 123 TVQGYPSVSSPQQGYYPGQASQPPGQGGQPKWQELGQGGQGYPTSLHSGGGGQGG 182

QY 135 YPPCS--RPEGGYGGGQ-----GORQYMP--PCRPG--TTGGG 167
DB 183 YPSSLQPGQGGQIQGGQGYPTSLQPGQCGQIQGGQGYPTSPQHPGQRQPGQG 242

QY 168 PRIGR-VRLTKAREYAAAG 184
DB 243 QIQGGQQLGQGRIGQG 260

RESULT 8
Q8LK18 PRELIMINARY; PRT; 704 AA.
AC Q8LK18;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE HMW glutenin subunit y.
GN GLU-15Y.
OS Aegilops speltoides (Goat grass).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Aegilops.
OX NCBI_TaxID=4573;
RN [1]
RP SEQUENCE FROM N.A.
RA De Buitos A.; Corredor M.; Jouve N.S.;
RT "Characterization of HMW glutenins in diploid species of the genus
RT Aegilops.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF513640; AAM47280.1; -.
DR InterPro; IPR001419; Glutenin.
DR Pfam; PF03157; Glutenin_hmw; 4.
DR PRINTS; PR00210; GLUTENIN.
SQ SEQUENCE 704 AA; 75932 MW; 1A4B73EC03C3F32C CRC64;

Query Match 24.7%; Score 283; DB 10; Length 704;
Best Local Similarity 34.8%; Pred. No. 1.7e-19;
Matches 77; Conservative 21; Mismatches 45; Indels 78; Gaps 9;

```

```

QY 16 ALVAVAVCOGVEVERQLRDLCQWQVQESPLDACRQVLDRLTGGGGGGVGPFRWGTGL 75
DB 14 ALVAVAVCOGVEVERQLRDLCQWQVQESPLDACRQVLDRLTGGGGGGVGPFRWGTGL 61
QY 76 MRCCQQLQDVSRRCRAAIRSMVRGYE-AMPPLEKGMWP-----WG--- 117
DB 62 MRCCQQLRDVSAKCRPVAIVQARQYEQTVVPKGGSFYPGETTPLOQLQGVIFWGTSS 121
QY 118 -----RQQPPPPQG-----GGGGGY----- 134
DB 122 QTQVGYPSVTSPPQGSYYPGQASQPPGQGGQPKWQELGQGGQGYPTSLQSGGQGG 181
QY 135 -YPPCS--RPEGGYGGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG 172
DB 182 GYPPSSLOQPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG 220

RESULT 9
Q40045 PRELIMINARY; PRT; 475 AA.
AC Q40045;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE D-hordein (Fragment).
GN HOR3.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=96204516; PubMed=8628236;
RA Sorensen M.B.; Muller M.; Skerritt J.; Simpson D.;
RT "Hordein promoter methylation and transcriptional activity in wild
RT type and mutant barley endosperm.";
RL Mol. Gen. Genet. 250:750-760(1996).
DR EMBL; X84368; CAA59104.1; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001419; Glutenin.
DR Pfam; PF03157; Glutenin_hmw; 3.
DR PRINTS; PR00210; GLUTENIN.
DR SMART; SM00499; AAI; 1.
FT CHAIN.
FT NON_TER 475 475
SQ SEQUENCE 475 AA; 50786 MW; 608B05E8E0F381B3 CRC64;

Query Match 24.6%; Score 282.5; DB 10; Length 475;
Best Local Similarity 36.2%; Pred. No. 1.2e-19;
Matches 79; Conservative 22; Mismatches 52; Indels 65; Gaps 11;

QY 1 MAKIAAAAAALCFALVAVAVCOGVE-----RQLRLDLCQWQVQESPLDACRQVLD 54
DB 1 MAKRLVLFVAVI--VALVALTTAREINGNNIFLDSRLSQCELEQLQESSLEACRQVVD 59
QY 55 RQLTGGGGGGVGPFRWGTGLMRCCQQLQDVSRRCRAAIRSMVRGYE----- 104
DB 59 QQL-----VGQLPWSGTGLQGCCQLRDVSPPCRVALSQVVRQYEQTEVPSKGS 110
QY 105 -----AMPPLEK-GWV-----PMGRQQQPPPGGGGGGGGY-----YPC 138
DB 111 FYPGGTAPPLQGGWGTGSKVKNYPDQTSQSQWQGGQGGQGGQGGQGGQGGQGGQGG 170
QY 139 S-----RPEGGYGGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG 166
DB 171 STFPQPGQGGQ---QPGQRQPSYPSATFPQPGQGGQ 205

RESULT 10
Q38767

```

Q38767 PRELIMINARY; PRT; 648 AA.
AC Q38767;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE high molecular weight glutenin subunit.
GN GLU-1-2.
OS Aegilops tauschii (Aegilops squarrosa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Aegilops.
OX NCBI_TaxID=37682;
RN [1]
RN SEQUENCE FROM N.A.
RA Mackie A.M., Sharp P.J., Lagudah E.S.;
RT "The nucleotide and derived amino acid sequence of a HMW glutenin gene
from Triticum tauschii and comparison with those from the D genome of
bread wheat.";
RL J. Cereal Sci. 0:0-0(1996).
RN [2]
RN SEQUENCE FROM N.A.
RA Mackie A.M., Lagudah E.S., Sharp P.J., Lafandra D.;
RT "Molecular and biochemical characterisation of HMW glutenin subunits
from T. tauschii and the D genome of hexaploid wheat.";
RL J. Cereal Sci. 0:0-0(1996).
DR EMBL; U39229; AAC15244.1; -.
DR InterPro; IPR003612; AAI.
DR Pfam; PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 648 AA; 69567 MW; B5D71332E520DF5E CRC64;

Query Match 24.6%; Score 282; DB 10; Length 648;
Best Local Similarity 37.1%; Pred. No. 1.9e-19;
Matches 79; Conservative 22; Mismatches 46; Indels 66; Gaps 11;

QY 1 MAKIAAAAAALCFALVAVAVCGEVRQLDLCWQVQSVQSPDLACRQVLDRLTGG 60
DB 1 MAKRLVLFRAVV--IALVALTTAEASRQ---LQERLQSSLEACRQVVDQL--- 51

QY 61 GGGGGVGFPRFWGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE--AMPPLEKGMWP--- 115
DB 52 -----AGRLPWSLTGLQMRCCQQLRDVSAKRSVAISQARQYEQTVVPPKGGSYFGETT 106
QY 116 -----WG-----RQ-----QQPPQGGGGGGG----- 133
DB 107 PLQLOQGIWFMTSSQTVOGYFPVTSRQGSYPGQASPPQGGQPGKQWQPGQGGQ 166
QY 134 YYPCSRPGEYGGGQGRQ--MYP-----PCRPGTTGG 162
DB 167 WYPTSLQPGGQQIGK-GQGGYPTSLQPG 198

RESULT 11
Q40054 PRELIMINARY; PRT; 707 AA.
AC Q40054;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE D hordein precursor.
GN HOR3.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RN SEQUENCE FROM N.A.
RA STRAIN=Haruna Nijo; TISSUE=Endosperm;
RA Hirota N., Kuroda H., Ito K.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; D82941; BAAL1642.1; -.
DR InterPro; IPR003612; AAI.
DR Pfam; PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
DR SMART; SM00499; AAI; 1.
KW Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 707 D HORDEIN.
SQ SEQUENCE 707 AA; 75108 MW; EB8873AF035D5789 CRC64;

Query Match 24.5%; Score 281.5; DB 10; Length 707;
Best Local Similarity 36.1%; Pred. No. 2.4e-19;
Matches 79; Conservative 22; Mismatches 51; Indels 67; Gaps 11;

QY 1 MAKIAAAAAALCFALVAVAVCGEVE-----RQLRLDLCWQVQSVQSPDLACRQVLD 54
DB 1 MAKRLVLFRAVV--IALVALTTAEIRNNGNIFLDSRQLQERLQSSLEACRQVVD 58
QY 55 ROLTGGGGGGVGFPRFWGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE----- 104
DB 59 QQL-----VGQLPWSLTGLQMRCCQQLRDVSAKRSVAISQARQYEQTVVPPKGG 110
QY 105 -----AMPPLEK-GMW-----PMGRQ-----QPPPGGGGGGGG 133
DB 111 FYPGGTAPPLQGGGWSVTKWYYPDQTSQQSWQGGYHQSIVTSQQPGQGGQSGYPG 170
QY 134 YYPCSRPGEYGGGQGRQ--MYP-----PCRPGTTGG 166
DB 171 STFP-QQPGGQ---QPGORQWSPSATFPQPGGQGG 205

RESULT 12
Q8H0L6 PRELIMINARY; PRT; 196 AA.
ID Q8H0L6
AC Q8H0L6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glutenin high molecular weight subunit (Fragment).
GN HMW-GS (DY TYPE).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Xiaoyan 6;
RA Fan S., Guo A., Luo D., Zhao H., Feng X.;
RT "Partial sequence of wheat HMW-GS (Dy type) gene from cultivar 'Xiaoyan 6'.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ308967; CAC84118.1; -.
FT NON_TER 196 196
SQ SEQUENCE 196 AA; 21587 MW; 4B48BE38F40A4D07 CRC64;

Query Match 24.2%; Score 277.5; DB 10; Length 196;
Best Local Similarity 36.7%; Pred. No. 1.4e-19;
Matches 76; Conservative 22; Mismatches 44; Indels 65; Gaps 10;

QY 1 MAKIAAAAAALCFALVAVAVCGEVRQLDLCWQVQSVQSPDLACRQVLDRLTGG 60
DB 1 MAKRLVLFRAVV--IALVALTTAEASRQ---LQERLQSSLEACRQVVDQL--- 51
QY 61 GGGGGVGFPRFWGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE--AMPPLEKGMWP--- 115
DB 52 -----AGRLPWSLTGLQMRCCQQLRDVSAKRSVAISQARQYEQTVVPPKGGSYFGETT 106
QY 116 -----WG-----RQ-----QQPPQGGGGGGG----- 133
DB 107 PLQLOQGIWFMTSSQTVOGYFPVTSRQGSYPGQASPPQGGQPGKQWQPGQGGQ 166

```
QY 134 YYPCS--RPGEYGYGGGQORQWYPP 158
Db 167 WYPTSLQPGGQQQIGKG--KQGYYP 191

RESULT 13
Q83V7 PRELIMINARY; PRT; 655 AA.
AC Q83V7;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE HMW glutenin subunit.
GN GLU-1UY.
OS Aegilops umbellulata.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Aegilops.
OX NCBI_TaxID=4491;
RN [1]_
RP SEQUENCE FROM N.A.
RA Liu Z., Yan Z., Wan Y., Liu K., Zheng Y., Wang D.;
RT "Analysis of HMW glutenin subunits and their coding sequences in two
RT diploid Aegilops species.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP476962; AAL82616.1; -.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF03157; Glutenin_hmw; 1.
DR Signal.
SQ SEQUENCE 655 AA; 70458 MW; 20FOA70E4D54FF2D CRC64;

Query Match 24.1%; Score 276.5; DB 10; Length 655;
Best Local Similarity 36.3%; Pred. No. 6.8e-19;
Matches 77; Conservative 21; Mismatches 45; Indels 69; Gaps 10;

QY 16 ALVAVAVCOGEVERQRLDLCQWQVQESPLDRCQVLDRLQLTGGGGGVGPFRRWGTGL 75
Db 14 ALVALTTAGEASRQ-----LQCERELQESLEACRQVVDQQL-----AGRLPWSTGL 61
QY 76 RMRCCQQLQDVSRCCRAIRSMVRGYE-AMPPLEKGMWP-----WGROQ 120
Db 62 QMRCCQQLRDVSAKRSVAVSQVQYEQTVPPKGSFYPGTTPLOQLQQGIFWGTSS 121
QY 118 -----RQ-----QQPPPGGGGGGGY-----YPPCS--RPGEYGY 146
Db 122 QTVOGYPSVTSRQSYYPGQASPPQGGQGGQVPPKGSFYPGTTPLOQLQQGIFWGTSS 121
QY 147 YGGGQORQYPPCRPGTTG-----GGPRIGR 172
Db 182 IGQG--QGGYPTSPQHTGQRQPPVGGQIQG 211

RESULT 14
Q94IK7 PRELIMINARY; PRT; 713 AA.
AC Q94IK7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE High molecular weight glutenin subunit y precursor.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Secale.
OX NCBI_TaxID=4550;
RN [1]_
RP SEQUENCE FROM N.A.
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314770; CAC40672.1; -.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF03157; Glutenin_hmw; 1.
DR Signal.
SQ SEQUENCE FROM N.A.
RA de Bustos A., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314781; CAC40682.1; -.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
DR SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT Y.
SQ SEQUENCE 713 AA; 76274 MW; 49EEEB182BBC19C0 CRC64;

Query Match 24.0%; Score 275; DB 10; Length 713;
Best Local Similarity 32.1%; Pred. No. 1.1e-18;
Matches 78; Conservative 21; Mismatches 46; Indels 98; Gaps 10;

QY 16 ALVAVAVCOGEVERQRLDLCQWQVQESPLDRCQVLDRLQLTGGGGGVGPFRRWGTGL 75
Db 14 ALVALTTAGEASRQ-----LQCERELQESLEACRQVVDQQL-----AGRLPWSTGL 61
QY 76 RMRCCQQLQDVSRCCRAIRSMVRGYE-AMPPLEKGMWP-----WGROQ 120
Db 62 QMRCCQQLRDVSAKRSVAVSQVQYEQTVPPKGSFYPGTTPLOQLQQGIFWGTSS 121
QY 121 Q-----PPQGG-----GGGGGGY----- 135
Db 122 QTVOGYPSVTSRQSYYPGQASPPQGGQGGQVPPKGSFYPGTTPLOQLQQGIFWGTSS 121
QY 136 --YPCS--RPGEY-----GYGGGQ-----RQYPPCRPGTTGGGPR 169
Db 182 GHYPASQQPGGQGGQGHYPASLQPPGQGGQGHYPASLQPPGQGGQGGQGGQGGQ 241
QY 170 IGR 172
Db 242 IQQ 244

RESULT 15
Q94IL4 PRELIMINARY; PRT; 737 AA.
AC Q94IL4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE High molecular weight glutenin subunit y precursor.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Secale.
OX NCBI_TaxID=4550;
RN [1]_
RP SEQUENCE FROM N.A.
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314770; CAC40672.1; -.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
DR SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT Y.
SQ SEQUENCE 713 AA; 76274 MW; 49EEEB182BBC19C0 CRC64;
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Imperial;
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314781; CAC40682.1; -.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
DR SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT Y.
SQ SEQUENCE 713 AA; 76274 MW; 49EEEB182BBC19C0 CRC64;

Query Match 24.0%; Score 275; DB 10; Length 713;
Best Local Similarity 32.1%; Pred. No. 1.1e-18;
Matches 78; Conservative 21; Mismatches 46; Indels 98; Gaps 10;

QY 16 ALVAVAVCOGEVERQRLDLCQWQVQESPLDRCQVLDRLQLTGGGGGVGPFRRWGTGL 75
Db 14 ALVALTTAGEASRQ-----LQCERELQESLEACRQVVDQQL-----AGRLPWSTGL 61
QY 76 RMRCCQQLQDVSRCCRAIRSMVRGYE-AMPPLEKGMWP-----WGROQ 120
Db 62 QMRCCQQLRDVSAKRSVAVSQVQYEQTVPPKGSFYPGTTPLOQLQQGIFWGTSS 121
QY 121 Q-----PPQGG-----GGGGGGY----- 135
Db 122 QTVOGYPSVTSRQSYYPGQASPPQGGQGGQVPPKGSFYPGTTPLOQLQQGIFWGTSS 121
QY 136 --YPCS--RPGEY-----GYGGGQ-----RQYPPCRPGTTGGGPR 169
Db 182 GHYPASQQPGGQGGQGHYPASLQPPGQGGQGHYPASLQPPGQGGQGGQGGQGGQ 241
QY 170 IGR 172
Db 242 IQQ 244

RESULT 15
Q94IL4 PRELIMINARY; PRT; 737 AA.
AC Q94IL4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE High molecular weight glutenin subunit y precursor.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Secale.
OX NCBI_TaxID=4550;
RN [1]_
RP SEQUENCE FROM N.A.
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314770; CAC40672.1; -.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
DR SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT Y.
SQ SEQUENCE 713 AA; 76274 MW; 49EEEB182BBC19C0 CRC64;
```

SQ SEQUENCE 737 AA; 79406 MW; 25E1BF9798290F06 CRC64;

Query Match 23.3%; Score 267.5; DB 10; Length 737;
 Best Local Similarity 31.8%; Pred. No. 6e-18;
 Matches 75; Conservative 18; Mismatches 44; Indels 99; Gaps 9;

QY 16 ALVAVACQGEVERQRLRDLCWQEVQESPLDACRQVLDRLQTLGGGGGGVGFPRWGTGL 75
 DB 14 ALVALTAAEGEASQ-----LQCEHELQESSLEACRQVVDQQL-----VGRLPWSTGL 61

QY 76 RMRCCQLODVSRCECAIRSMVRGYER-AMPPLEKHWHP-----WGRQQ 120
 DB 62 HMRCCQQLRDVSAKCRHVAVSQVARYEQETVPPKGGSFYSETPLQQLQOGIFWGTSS 121

QY 121 Q-----PPPOGG-----GGGQGGY----- 135
 DB 122 QTVQGYPSVTSFQQGSYYPGRASPOQPGQGOQPGKWQEPGQGOQGYFTSLQQPGQGOQ 181

QY 136 -----YPCS--RPEGYGYGGGQORQMYPPCRPPTTGGGPRIQR 172
 DB 182 EHYPASQQQPGQGGQGHYPASLQPPGQGOQTGPQGIQO-----PGGQGOIQG 229

Search completed: November 29, 2003, 17:06:33
 Job time : 61 secs

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	130.5	11.4	157	3	US-08-706-391B-7	Sequence 7, Appli
2	115	10.0	308	4	US-09-252-991A-27249	Sequence 27249, A
3	114.5	10.0	410	4	US-09-252-991A-25812	Sequence 25812, A
4	109.5	9.5	426	1	US-07-918-023-2	Sequence 2, Appli
5	103	9.0	476	4	US-09-252-991A-17887	Sequence 17887, A
6	101.5	8.8	863	4	US-09-252-991A-19574	Sequence 19574, A
7	100.5	8.8	535	4	US-09-252-991A-28410	Sequence 28410, A
8	100	8.7	209	4	US-09-252-991A-30648	Sequence 30648, A
9	100	8.7	656	2	US-08-343-433B-2	Sequence 2, Appli
10	100	8.7	656	3	US-09-214-564A-4	Sequence 4, Appli
11	99	8.6	238	4	US-09-252-991A-17915	Sequence 17915, A
12	98.5	8.6	124	1	US-08-216-846-2	Sequence 2, Appli
13	98.5	8.6	124	1	US-08-782-177-3	Sequence 3, Appli
14	98.5	8.6	569	4	US-09-252-991A-24513	Sequence 24513, A
15	98.5	8.6	639	4	US-09-252-991A-28453	Sequence 28453, A
16	97.5	8.5	393	4	US-09-252-991A-30202	Sequence 30202, A
17	97	8.4	148	1	US-08-207-904-15	Sequence 15, Appl
18	96.5	8.4	188	4	US-09-252-991A-24358	Sequence 24358, A
19	96.5	8.4	493	3	US-08-556-978B-9	Sequence 59, Appl
20	96.5	8.4	494	4	US-09-252-991A-32227	Sequence 32227, A
21	96.5	8.4	529	3	US-09-247-806-2	Sequence 2, Appli
22	96	8.4	456	4	US-09-252-991A-19417	Sequence 19417, A
23	96	8.4	1958	1	US-07-945-283-2	Sequence 2, Appli
24	95.5	8.3	397	4	US-09-252-991A-25695	Sequence 25695, A
25	95.5	8.3	656	4	US-09-252-991A-22270	Sequence 22270, A
26	94.5	8.2	425	4	US-09-252-991A-31834	Sequence 31834, A
27	94.5	8.2	859	1	US-08-395-580-2	Sequence 2, Appli

QY 81 QQLQDVSRCCAAIRSMVRGYEE--AMPLEKGMWPGQQQPPQ-----GGGGGG 132
DB 32 QQLRDISPECHPVVAVAGYEQQIVVPP--KGYTFYGETTTPQOQRIFWGIPALL 89
QY 133 GYVYP---CSRPGEGYGGGQGRQVMP-----PCRPTT---GGGPRIGRVLTKAREVA 182
DB 90 KRYPSVTCFQ-----QVSYFPGQASQPSRSSSVHVSVEHQASLKVAKAQQOLA 138
QY 183 AGLPMCMRLSEPOECSIFSGD 204
DB 139 AQLPAMCRL-----EGGD 151

RESULT 2

US-09-252-991A-27249
; Sequence 27249, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27249
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27249

Query Match 10.0%; Score 115; DB 4; Length 308;
Best Local Similarity 25.0%; Pred. No. 0.0012;
Matches 64; Conservative 13; Mismatches 69; Indels 110; Gaps 14;
QY 6 AAAAAALCFAALVAVACQEVERQRLDLQCVQEVQESPLDACRQVLDRLQLTGGGGGG 65
DB 2 SAAARTASGAVATAPAG-----AGPGRGG 28
QY 66 VGFPRWG-----TGLRM-----RCCQQLQDVSRCCAAIRSMVRGY 102
DB 29 SGPGRGRSIPAPRRGPRLARPSRGRTRAPGRCCSGW--STGSPMAPVRSACRG 86
QY 103 EAMPLEKGMW-----WGROQP-----PPQGGG-----CGGY 134
DB 87 SR---PVRS--WGPAPRAGGYGRRRGRPRRSVPPTVASVAGPSRSGAGPARRAAGPGA- 140
QY 135 YVPCSRPGEGYVG-----CGGQRMVPPC-----RP-GTGGGPRIGRVLTKARE 180
DB 141 -APCPGGRSGAARRHRGRGRSRRWCPCAGSARRPSGCTGCRPLPR---RAGRR 196
QY 181 YAAGLPMCMRLSEPOE 196
DB 197 AAGRPCCPPGAPGE 212

RESULT 3

US-09-252-991A-25812
; Sequence 25812, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25812
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25812

Query Match 10.0%; Score 114.5; DB 4; Length 410;
Best Local Similarity 36.2%; Pred. No. 0.0018;
Matches 47; Conservative 10; Mismatches 54; Indels 19; Gaps 8;
QY 61 GGGGG-----VGFPR-WGTGLRMCCQQLQDVSRCCAAIRSMVRGYEAMPLEKGMW 115
DB 81 GAGGRCRRAEPWRTQGTGTRRAAPAA-----VARACRCLPRQRIACVARTAPGGERRAGP 136
QY 116 WGRQQPPPPQGGGGGGYVYPCSRPGEGY--YGGGQRMVPPCPRGTTGGGPRIGRVR 174
DB 137 -GRPR--SAGGAPRGGGARLAPPPAGPGRGTGQAGRR-----CRPGGTPGHR--RPR 187
QY 175 LTKAREYAAAG 184
DB 188 PEPRGDGAAG 197

RESULT 4

US-07-918-023-2
; Sequence 2, Application US/07918023
; Patent No. 5427934
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Thomas
; APPLICANT: Robins, Karen
; APPLICANT: Birch, Olwen
; APPLICANT: Bohlen, Elisabeth
; TITLE OF INVENTION: GENETIC ENGINEERING PROCESS FOR THE
; TITLE OF INVENTION: PRODUCTION OF S-(+)-2,2-DIMETHYLCYCLOPROPANECARBOXAMIDE
; TITLE OF INVENTION: BY MICROORGANISMS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fisher, Christen & Sabol
; STREET: 2000 M Street, N.W., Suite 590
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/918,023
; FILING DATE: 19920724
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2247/91
; FILING DATE: 26-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Marsh, Virgil H.
; REGISTRATION NUMBER: 23,083
; REFERENCE/DOCKET NUMBER: LP 1521
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2000
; TELEFAX: (202) 659-2015
; TELEX: 248748
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 426 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; PUBLICATION INFORMATION:

DOCUMENT NUMBER: IE 9224406
FILING DATE: 23-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426
PUBLICATION INFORMATION:
DOCUMENT NUMBER: JP 198717/92
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426
PUBLICATION INFORMATION:
DOCUMENT NUMBER: PL P 295408
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426
PUBLICATION INFORMATION:
DOCUMENT NUMBER: RO 92-01033
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426
PUBLICATION INFORMATION:
DOCUMENT NUMBER: SU UNKNOWN
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426
PUBLICATION INFORMATION:
DOCUMENT NUMBER: CS PV2323-92
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426
PUBLICATION INFORMATION:
DOCUMENT NUMBER: HU P9202439
FILING DATE: 24-JUL-1992
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426
US-07-918-023-2

Query Match 9.5%; Score 109.5; DB 1; Length 426;
Best Local Similarity 22.8%; Pred. No. 0.006;
Matches 64; Conservative 16; Mismatches 70; Indels 131; Gaps 12;
QY 24 QGEVERQRRLDLCQWQVQESPLD--ACQVLDRLQLTGGGGGGV----- 66
DB 123 QAAADRRRLRPPSRDCHPQPLERPMARGVLQRLGRGHGGAVLRIAGHGRHLPLS 182
QY 67 -----GPFWRGTGLMRCCQQLQDVSRRCRCRAIRSM 98
DB 183 IGRQRHHGAQHLGQCEPPRLRTGRVPGPHR-----PDGAQCCRC---RSH 226
QY 99 VRGYEAMP-----PLEK-----CWMP---WGRQQQPPPGGGGGGG 133
DB 227 ARGHRRGGPAGPYQPVQARLSGHDHARILRPAGHPGAMGTGRGRCPLPGGGAGPGG 286
QY 134 YYPFCSRPGEGYGGGQGORQMP--PCRPTTGGG----- 167
DB 287 -----GAAPGQORAGGGLSRCHPG--GGGLAGAVRGGRRRGARRHVPCTARGL 332
QY 168 -PRIGRVRLTAKAREYAAGLPMCMCR---LSEPOECISFSG 203
DB 333 WPARVRDRPGAGPVHRRLPAAAAAPRGLHGPACTLRAGG 373

RESULT 5
US-09-252-991A-17887
; Sequence 17887, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17887
; LENGTH: 476
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17887
Query Match 9.0%; Score 103; DB 4; Length 476;
Best Local Similarity 25.9%; Pred. No. 0.03;
Matches 60; Conservative 17; Mismatches 79; Indels 76; Gaps 10;
QY 2 AKIAAAAAALCFAALVAVAVQCGE-VERQRLRLDLCQWQVQESPLDCAKQVLDRLQLTGG 60
DB 137 ASRSADAARAVPAIAGASAAERQGRPAGDQRLR-----PADR-----DRLADGR 180
QY 61 GGGGGVGFPRMGTGLMRCCQQLQDVSRRCRCRAIRSMVRGYEEMPPPL-----EKG 112
DB 181 RGGAG-----QAGLHLRAWQ---GRADRPVAATAAAGRGGAATARIAGAAGSADR 231
QY 113 WWPWGRQQPPPGGGGGGGYYPFCSRPE-----GYGYGGG----- 151
DB 232 -----GAQGARPPQGRAGGQGRTHPGARGDGLARRGEAGLHGYGAGATAGPRLGAG 287
QY 152 -----QROMYPPCRPTTGGG--PRIGRVRLTAKAREYAAGLP 186
DB 288 LAAARQGLDGLALNLDRLTRPTCRPGSGADVEPAAGLADDTFARVAVRGDP 339

RESULT 6
US-09-252-991A-19574
; Sequence 19574, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19574
; LENGTH: 863
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19574

Query Match 8.8%; Score 101.5; DB 4; Length 863;
Best Local Similarity 29.5%; Pred. No. 0.082;
Matches 39; Conservative 14; Mismatches 48; Indels 31; Gaps 7;
QY 59 GGGGGVGFPRMGTGLMRCCQQLQDVSRRCRCRAIRSMVRGYEEMPPLEKQWMPW-- 116
DB 155 GAAGPGAVGRNRSTGIRRL-----GTRRTAVRPTLRRCRTSVRPVH---LPWLS 202
QY 117 GRQQPPPGGGGGGGYYPFCSRPEGYGGGQGORQMPYPCPTTGGGPRIGRVRT 176
DB 203 GR-ADPPYRPHRGQACPAFTGRPGG-----RGERR---PVVRTPGAPRL----- 247
QY 177 KAREYAAGLPM 188
DB 248 --RRPAGGVRL 257

RESULT 7
US-09-252-991A-28410
; Sequence 28410, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 28410
;; LENGTH: 535
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28410

Query Match 8.8%; Score 100.5; DB 4; Length 535;
Best Local Similarity 26.7%; Pred. No. 0.06;
Matches 62; Conservative 12; Mismatches 71; Indels 87; Gaps 12;

QY 25 GEVERORLRLQWQVQSPDLQACRQVLDRLQTGG-----GGGGVGFRRGTGL 75
DB 167 GQDRPRADRAQ-----PAAACR---RHAGNPRGVLLGEPAGPRPVGEHPWHPGL 214
QY 76 RRECCOQLQDVS-----ECRCAIRSMVRYEAMPPEKGMWPGRQ-----119
DB 215 RAG-----QHVGRHRAPELPAERPPAAQRDALRSRAA---QRGHRQRLGHGRRRG 265
QY 120 ----QOPP----PQGGGGGGGY--YPCSRPG-----EGYGGGQGOR 153
DB 266 DRDRQLPHRGPRPGQACAGRWPGAPHQPRRQWPHLRQRLHRRHGLGHAGRR 325
QY 154 QMYPP--CRPTTGGPRIGRVLTKAREYAAGLPMWMLSEPEQCSIFSGG 203
DB 326 QRAPPGLRPGDQGHWAHRLVQSRGAAG-----QAFAGG 364

RESULT 8

US-09-252-991A-30648
Sequence 30648, Application US/09252991A
Patent No. 6551795

;; GENERAL INFORMATION:
;; APPLICANT: Marc J. Rubenfield et al.

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 30648
;; LENGTH: 209
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30648

Query Match 8.7%; Score 100; DB 4; Length 209;
Best Local Similarity 27.8%; Pred. No. 0.023;
Matches 52; Conservative 12; Mismatches 70; Indels 58; Gaps 11;

QY 48 ACRQVLDRLQTLGGGGGG-----VGPFRWGTGLRMRCQQLQDVSRRCRCAAIRSMVR 100
DB 17 AQPAPRRHVAAGGDDQMAQLPAYLAGAHPAARG--QRAVRLRLDRPSAQRAGGLHT---71
QY 101 GVEEAMPP--LEKWWPWGQQQPPPGG-----GGGGGYYPYC-----SRP 141
DB 72 -----WPPGLLEPG--PGGRATHPDAPGHRTAGGNRAGDPGRGYRYGARPHRRHRHP 124
QY 142 GEGY-----GYGQGORQMYPCRTGTTGGPR--IGVR---LTKAREYAA 183
DB 125 GPCQRTAAVDCGGRPRGRGAGRAQSPAFAFERRGAPQPLVLRSMPTLPDDRQSA 184
QY 184 GLPMMCLSEPPQ 195

DB 185 VYP--CORSAP 194

RESULT 9

US-08-343-443B-2
Sequence 2, Application US/08343443B
Patent No. 5968734

;; GENERAL INFORMATION:

;; APPLICANT: Aurias, Alain
;; APPLICANT: Delattre, Olivier
;; APPLICANT: Desmaze, Chantal
;; APPLICANT: Melot, Thomas
;; APPLICANT: Peter, Martine
;; APPLICANT: Plooungastel, Beatrice
;; APPLICANT: Thomas, Gilles
;; APPLICANT: Zucman, Jessica

;; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
;; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
;; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
;; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
;; TITLE OF INVENTION: TRANSLOCATIONS
;; NUMBER OF SEQUENCES: 129
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Weiser & Associates
;; STREET: 230 South Fifteenth Street
;; CITY: Philadelphia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19102

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: AEDIT 1.0 DOS text editor
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/343,443B
;; FILING DATE: 18-NOV-1994
;; CLASSIFICATION: 514

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/FR93/00494
;; FILING DATE: 19-MAY-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: FR 92/06123
;; FILING DATE: 20-MAY-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weiser, Gerard J.
;; REGISTRATION NUMBER: 19,763
;; REFERENCE/DOCKET NUMBER: 989.6121P

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-875-8383
;; TELEFAX: 215-875-8394

;; INFORMATION FOR SEQ ID NO: 2:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 656 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

US-08-343-443B-2

Query Match 8.7%; Score 100; DB 2; Length 656;
Best Local Similarity 38.4%; Pred. No. 0.084;
Matches 28; Conservative 6; Mismatches 25; Indels 14; Gaps 3;

QY 99 VRYEAMPPLEKGMWPGRQOQPPPGGGGGGGYYPYPCSPGEGYGYGGGQROMYPP 158
DB 454 MRG---GLPPR-----GRGMPPLRGGPGGPGGPMGRMG-----GCGDRGGGFP 499

QY 159 CRPGTTGGGPRIG 171

DB 500 RGRGRGRGNPGG 512

RESULT 10

US-09-214-564A-4
; Sequence 4, Application US/09214564A
; Patent No. 6150515
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip A.
; APPLICANT: Zhou, Qiang
; TITLE OF INVENTION: Tat-SF: Cofactor For Stimulation Of Transcriptional
; FILE REFERENCE: M0656/7042
; CURRENT APPLICATION NUMBER: US/09/214,564A
; PRIOR FILING DATE: 1999-08-18
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/033,152
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: PCT/US97/11713
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-564A-4

Query Match 8.7%; Score 100; DB 3; Length 656;
Best Local Similarity 38.4%; Pred. No. 0.084;
Matches 28; Conservative 6; Mismatches 25; Indels 14; Gaps 3;

QY 99 VRGYEAMPLEKGNWPMGRQPPQGGGGGGYPCSRPEGYGGQGGQRMYP 158
DB 454 MRG---GLPPE-----GRGMPEPLRGPGGPGGPMGRMG-----GRGGDRGGFPF 499

QY 159 CRPGTTGGGPRIG 171
DB 500 RGRGRSGRGNPSGG 512

RESULT 11
US-09-252-991A-17915
; Sequence 17915, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17915
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17915

Query Match 8.6%; Score 99; DB 4; Length 238;
Best Local Similarity 28.1%; Pred. No. 0.033;
Matches 39; Conservative 8; Mismatches 56; Indels 36; Gaps 7;

QY 49 CRVLDRLQTLG-----GGGGGGVPRFMGTGLRMRCQQQLQDVSRRCRAIRSMV 102
DB 114 CRTVARRATGRRGASSGGAAGSAPRATAPGIRRR-----SVPARCGCVA-----AGY 162

QY 103 EAMPLEKGNWPMGRQPP-----PQGGGGGGYPCSRPEGYG---YGGGGQRMYP 156
DB 163 WRSAREPLA---WPRGRGSAFVRGPRRRSGSAPGTAPRRRTGTGRAIRHDSGRHVC 219

QY 157 PP-----CRPGTTG 165

DB 220 VPASACGARCGAARPGACG 238

RESULT 12

US-08-216-846-2
; Sequence 2, Application US/08216846
; Patent No. 5444046
; GENERAL INFORMATION:
; APPLICANT: MIYAZAKI, TOSHIYUKI
; APPLICANT: MORIMOTO, TOSHIHISA
; APPLICANT: MURAYAMA, RYUJI
; APPLICANT: MATSUBARA, HIROSHI
; TITLE OF INVENTION: AMYLASE INHIBITORS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,846
; FILING DATE: 24-MAR-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 91881/1993
; FILING DATE: 29-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 148423/1993
; FILING DATE: 28-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5444046man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1327-029-0X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-846-2

Query Match 8.6%; Score 98.5; DB 1; Length 124;
Best Local Similarity 27.6%; Pred. No. 0.018;
Matches 43; Conservative 8; Mismatches 38; Indels 67; Gaps 7;

QY 39 QEVQSPDLACRQVLDRLQTLTGGGGGGVPRFMGTGLRMRCQQQLQDVSRRCRAIRSM 98
DB 10 QAFQVPALPACRPLRLRLQCNQSGVPEAV-----LR-DCCQLAHISEWCRCGALYSM 60
QY 99 VRG-YEEAMPLEKGNWPMGRQPPQGGGGGGYPCSRPEGYGGQGGQRMYP 157
DB 61 LDSMYKE-----HGAQEGQAG----- 76
QY 158 PCRPGTTGGPRIGR--VELTKAREYVAGLPMWCR 191
DB 77 -----TGAFPRRREVVKLT-----AASITAVCR 101

RESULT 13

US-08-782-177-3

```

: Sequence 3, Application US/08782177
: Patent No. 5789380
: GENERAL INFORMATION:
: APPLICANT: MIYAZAKI, TOSHIYUKI
: APPLICANT: MORIMOTO, TOSHIHISA
: APPLICANT: MURAYAMA, RYUJI
: APPLICANT: TAKASE, SACHIKO
: APPLICANT: GODA, TOSHINO
: TITLE OF INVENTION: AGENTS FOR INHIBITING ACCUMULATION OF
: VISCERAL FAT
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSES: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
: ADDRESSES: P.C.
: STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
: CITY: ARLINGTON
: STATE: VA
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/782.177
: FILING DATE: 14-JAN-1997
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 23445/1996
: FILING DATE: 18-JAN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: OBLON, NORMAN F.
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 1327-055-0X
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-413-3000
: TELEFAX: 703-413-2220
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 124 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-782-177-3

```

Query Match	8.6%	Score 98.5;	DB 1;	Length 124;
Best Local Similarity	27.6%;	Pred. No. 0.018;		
Matches	43;	Conservative	8;	Mismatches 38; Indels 67; Gaps 7;
QY	39	QEVQESPLDACRQVLDRLQLTGGGGGGVGFRWGTGLMRCCOOLQDVSECRCAATISM	98	
Db	10	QAFQVPALPACRPLRLQCNQSGVPEAV-----LR-DCCOQLAHISEWCRGALYSM	60	
QY	99	YEEAMPPELEKGMWPMCRQPPPPQGGGGGGGYYYPCSRPGEVGYGGGQRQWYP	157	
Db	61	LDSMYXE-----HGAQEGQAG-----	76	
QY	158	PCRPGTTGGPRIGR--VRLTKAREYAAGLPMCMRL	191	
Db	77	-----TGAFPRCRREVKLT-----AASITAVCRL	101	

RESULT 14
US-09-252-991A-24513
; Sequence 24513, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

```

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24513
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24513

```

Query Match	8.6%;	Score 98.5;	DB 4;	Length 569;
Best Local Similarity	27.7%;	Pred. No. 0.1;		
Matches 53;	Conservative 12;	Mismatches 65;	Indels 61;	Gaps 11;
QY	48	ACR-----QVLRLQLTG-----GG-----GGGGVGPFRWG-----TGLRMRCQ- 81		
Db	333	ACRPRFGATGDRRLRGATADQPDGRRQRPSGGHYRRLLTAQRRSPAGRLRLCRL 392		
QY	82	-----QLQDVSRECR-----CAAIRSMVTGYEAM--PPELKGWMPWGRRQQPPPOGGGGGQG 132		
Db	393	SGQQLRFAAGQARRAWCGLSRVAFIPCLAIRPPGGRGEQ- GSAAQPSAASGGTGQP 451		
QY	133	GVYVPCSR--PPEGYG-----YGGGQRQWYPPCRPGTTGGGP 168		
Db	452	G-RHPLRRPPGDTAGAYPGEPSAAPRRVPARRGRACAGRGATRPAAVCAQRGANGGP 510		
QY	169	RIGRVLTKAR 179		
Db	511	GSGRWRTGACR 521		

```

RESULT 15
US-09-252-991A-28453
, Sequence 28453, Application US/09252991A
, Patent No. 6551795
, GENERAL INFORMATION:
, APPLICANT: Marc J. Rubenfield et al.
, TITLE OF INVENTION: NUCLEIC ACID AND
, TITLE OF INVENTION: AERUGINOSA FOR DIA
, FILE REFERENCE: 107196.136
, CURRENT APPLICATION NUMBER: US/09/252,9
, CURRENT FILING DATE: 1999-02-18
, PRIOR APPLICATION NUMBER: US 60/074,78
, PRIOR FILING DATE: 1998-02-18
, PRIOR APPLICATION NUMBER: US 60/094,19
, PRIOR FILING DATE: 1998-07-27
, NUMBER OF SEQ ID NOS: 33142
, SEQ ID NO 28453
, LENGTH: 639
, TYPE: prt
, ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28453

```

Query Match	8.6%;	Score	98.5;	DB	4;	Length	639;
Best Local Similarity	26.7%;	Pred. No.	0.11;				
Matches	58;	Conservative	17;	Mismatches	73;	Indels	69;
Gaps	13;						
QY	25	GEVEROR--LRDL--QCQWEQVESPLDCAQVLDQLTGGGG--	GGVGVPFRW	71			
Db	117	GLVEHQAGRLRLARRHVLRGGQHPLPALHRAAGR--GGAGVHPQAAEGTGLHPVRR	173				
QY	72	GTGLRM-----RCCQQLQDVSECRCAATRSVMR-----	100				
Db	174	ARGHRVDPGADHRCAPANESHAGQLRHLRLQRRRRRGDGRPGTARRYPAA	233				
QY	101	GVEEAMPP-LEKGWPMWQQQPPPGGG--GGGGGY--YPCSRPBGEGYGGGGORQM	155				
Db	234	GTGNNPLQWDRGVWMLRGAQPPGGGRLAGQAQALRRLPGGRPGD-----PRRL	286				
QY	156	YPPCRDGTTCGGPRIGRV--RITKAREYAAGLPMCMR	190				

[illegible]

Search completed: November 29, 2003, 17:07:34

Job time : 25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2003, 17:06:40 ; Search time 46 Seconds
(without alignments)
825.985 Million cell updates/sec

Title: US-10-053-410-4
Perfect score: 1148
Sequence: 1 MAKIAAAAAALCFALNAV.....MMCRILSEPOECIFSGGDQY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673684 seqs, 184443283 residues

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117	10.2	169	12	US-10-165-289A-4
2	106.5	9.3	1145	10	US-09-866-557A-5
3	106.5	9.3	1145	15	US-10-055-797-5
4	105	9.1	168	12	US-10-165-289A-2
5	105	9.1	277	11	US-09-847-102A-73
6	105	9.1	694	11	US-09-847-102A-56
7	105	9.1	694	12	US-10-285-976-53
8	105	9.1	694	15	US-10-205-823-144
9	101	8.8	591	15	US-10-233-553-23
10	101	8.8	894	15	US-10-233-553-11
11	100.5	8.8	544	12	US-10-017-161-670
12	97.5	8.5	302	14	US-10-029-180-18
13	97.5	8.5	623	15	US-10-156-761-11120
14	96.5	8.4	529	9	US-09-861-597-2
15	96	8.4	2331	15	US-10-184-644-59

16	96	8.4	2331	15	US-10-184-634-59	Sequence 59, Appl
17	94.5	8.2	520	12	US-10-168-097A-36	Sequence 36, Appl
18	94.5	8.2	520	12	US-10-239-431A-34	Sequence 34, Appl
19	94.5	8.2	859	15	US-10-153-668-324	Sequence 324, Appl
20	94.5	8.2	892	15	US-10-153-668-438	Sequence 438, Appl
21	94	8.2	485	10	US-09-925-300-1664	Sequence 1664, Appl
22	94	8.2	775	15	US-10-156-761-12824	Sequence 12824, A
23	93.5	8.1	132	9	US-09-864-761-43644	Sequence 43644, A
24	93.5	8.1	166	15	US-10-157-031-80	Sequence 80, Appl
25	93.5	8.1	238	15	US-10-204-887-116	Sequence 116, Appl
26	92.5	8.1	599	11	US-09-847-102A-37	Sequence 37, Appl
27	92.5	8.1	755	10	US-09-919-497-57	Sequence 57, Appl
28	92	8.0	274	11	US-09-847-102A-74	Sequence 74, Appl
29	92	8.0	682	15	US-10-152-548-16	Sequence 16, Appl
30	92	8.0	685	11	US-09-847-102A-57	Sequence 57, Appl
31	92	8.0	685	15	US-10-146-474-14	Sequence 14, Appl
32	91.5	8.0	148	12	US-10-447-541-6	Sequence 6, Appl
33	91.5	8.0	253	12	US-10-304-630-7	Sequence 7, Appl
34	91.5	8.0	253	12	US-10-304-630-9	Sequence 9, Appl
35	91.5	8.0	253	12	US-10-304-630-10	Sequence 10, Appl
36	91.5	8.0	253	12	US-10-304-630-11	Sequence 11, Appl
37	91.5	8.0	253	12	US-10-304-630-12	Sequence 12, Appl
38	91.5	8.0	253	12	US-10-304-630-14	Sequence 14, Appl
39	90.5	7.9	276	12	US-10-017-161-2268	Sequence 2268, Ap
40	90.5	7.9	481	15	US-10-156-761-12906	Sequence 12906, A
41	90.5	7.9	752	15	US-10-156-761-9350	Sequence 9350, Ap
42	90.5	7.9	1058	15	US-10-174-363-38	Sequence 38, Appl
43	90.5	7.9	3060	15	US-10-184-644-337	Sequence 337, Appl
44	90.5	7.9	3060	15	US-10-184-634-337	Sequence 337, Appl
45	90	7.8	378	9	US-09-849-967A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-10-165-289A-4
; Sequence 4, Application US/10165289A
; Publication No. US2003015917A1
; GENERAL INFORMATION:
; APPLICANT: QIU, Xiao
; APPLICANT: TRUSKA, Martin
; APPLICANT: HU, Zhiyuan
; TITLE OF INVENTION: Flax (Linum usitatissimum L.)
; TITLE OF INVENTION: Seed-Specific Promoters
; FILE REFERENCE: BNZ-005
; CURRENT APPLICATION NUMBER: US/10/165,289A
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/295823
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Linum usitatissimum
US-10-165-289A-4

Query Match	10.2%	Score 117;	DB 12;	Length 169;
Best Local Similarity	23.9%	Pred. No. 0.0097;		
Matches	56;	Conservative 24;	Mismatches 56;	Indels 98; Gaps 13;
QY	1	MAKIAAAAAALCFALNAVAVCQGEVERQRRLDLCQWQESPLDACRQ----	VLDROL 57	
DB	1	MAKMSLAATAVAFILFIVV-----	DASVRTTVIIDEE- 33	
QY	58	TGGGGGGGGVGPFRWGTGLMRCCQOL--	QDVSRRCRCALRSMVRGVEAMPLEKGNP 115	
DB	34	TNQGRRGGGQ-----	GQGGQSCBQIQQDFLRSCQ-----	QFM 68
QY	116	WGRQQQPPPGGGGGGGYYPFCRPGEGYGGGQGRQMYPPC-----	RPGTGGG- 167	
DB	69	WEKVQR-----	GGRSHYV-----	NQGRGGGQSQYFDSCCDDLKQLSTGCTCRGL 113

```
QY 168 -PRIGVR-----LTKAREYAAAGLPMWCRSLRQECISIFSGDQ 205
Db 114 ERAIGQMRBIEIQGGQEQVQRWTAQAKTAKLPQOCR-TQPSQCO-FQGGQO 165

RESULT 2
US-09-866-557A-5
; Sequence 5, Application US/09866557A
; Patent No. US20020162126A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, S.
; APPLICANT: Hammond, G.
; APPLICANT: Beach, D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RNA INTERFERENCE
; FILE REFERENCE: GNCA-P03-007
; CURRENT APPLICATION NUMBER: US/09/866,557A
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: 60/189,739
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/243,097
; PRIOR FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1145
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-866-557A-5

Query Match 9.3%; Score 106.5; DB 10; Length 1145;
Best Local Similarity 25.9%; Pred. No. 0.54;
Matches 45; Conservative 25; Mismatches 71; Indels 33; Gaps 6;

QY 30 QRLRLDQCWQEVQESPLDACRQVLDRLTGCGGGGGVGPFRWG-----TGLRM 77
Db 47 QQLQPPQ--QQQQQPHQQQQSSRQPSSTSSGSRASGFQGGGQQKSDAEGWTAQKK 104
QY 78 RCCQQLQDVSRRCRA-----AIRSMVRGYEEAMPPEKGMWPGRQ-----QOPPP--- 124
Db 105 QGKQVQGWTKQGQGGHQQGQCGDGGYQQRPPGQQGQGGHQQGQCGEGGYQQRPPGQ 164
QY 125 -----QGGGGGGGGY--PCSRPGEYGYGGGQGMYPGCRPRTTGGGPRIGR 172
Db 165 QGGHQQGQCGEGYQQRPSGGQQGQGGHQQGQCGEGGYQQRPPGQQGQGGHQQGR 218

RESULT 3
US-10-055-797-5
; Sequence 5, Application US/10055797
; Publication No. US2003008471A1
; GENERAL INFORMATION:
; APPLICANT: Hammond et al.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RNA INTERFERENCE
; FILE REFERENCE: GNCA-P03-007
; CURRENT APPLICATION NUMBER: US/10/055,797
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: 60/189,739
; PRIOR FILING DATE: PCT/US01/08435
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1145
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-055-797-5

Query Match 9.3%; Score 106.5; DB 15; Length 1145;
Best Local Similarity 25.9%; Pred. No. 0.54;
Matches 45; Conservative 25; Mismatches 71; Indels 33; Gaps 6;

QY 30 QRLRLDQCWQEVQESPLDACRQVLDRLTGCGGGGGVGPFRWG-----TGLRM 77
```

```
Db 47 QQLQPPQ--QQQQQPHQQQQSSRQPSSTSSGSRASGFQGGGQQKSDAEGWTAQKK 104
QY 78 RCCQQLQDVSRRCRA-----AIRSMVRGYEEAMPPEKGMWPGRQ-----QOPPP--- 124
Db 105 QGKQVQGWTKQGQGGHQQGQCGDGGYQQRPPGQQGQGGHQQGQCGEGGYQQRPPGQ 164
QY 125 -----QGGGGGGGGY--PCSRPGEYGYGGGQGMYPGCRPRTTGGGPRIGR 172
Db 165 QGGHQQGQCGEGGYQQRPSGGQQGQGGHQQGQCGEGGYQQRPPGQQGQGGHQQGR 218

RESULT 4
US-10-165-289A-2
; Sequence 2, Application US/10165289A
; Publication No. US20030159174A1
; GENERAL INFORMATION:
; APPLICANT: QIU, Xiao
; APPLICANT: TRUSKA, Martin
; APPLICANT: HU, Zhiyuan
; TITLE OF INVENTION: Flax (Linum usitatissimum L.)
; FILE REFERENCE: BNZ-005
; CURRENT APPLICATION NUMBER: US/10/165,289A
; CURRENT FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/295823
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Linum usitatissimum
US-10-165-289A-2

Query Match 9.1%; Score 105; DB 12; Length 168;
Best Local Similarity 26.5%; Pred. No. 0.11;
Matches 39; Conservative 21; Mismatches 45; Indels 42; Gaps 7;

QY 1 MAKIAAAAAALCPAALVAVAVC-----QGEVERQRLRLDQCWQEVQ 43
Db 1 MAKLAALAAVATAFLIIVVDASVRTTVIDEDTNQGGGGGGGQQQ---QCEKQIOE 56
QY 44 SP-LDACROVL-DRLTGG-----GGGGGVGPFRWGTGLMRCCQQLQDVSRRCRA 94
Db 57 QDYLRCQCFLEKVKQGRSRYNNGRGGGQSQHFDS-----CCDDLKQLRSECTC--- 109
QY 95 IRSMVRGYEEAMPPEKGMWPGRQ 121
Db 110 -----RGLERAIGMRQDIQQGQGGQ 131

RESULT 5
US-09-847-102A-73
; Sequence 73, Application US/09847102A
; Publication No. US2003004409A1
; GENERAL INFORMATION:
; APPLICANT: University of California
; APPLICANT: Carson, Dennis A.
; APPLICANT: Cort, Maribat
; APPLICANT: Rhee, Chae-Seo
; APPLICANT: Lorenzo, Leoni M.
; APPLICANT: Malini, Sen
; TITLE OF INVENTION: IMMUNOLOGIC COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 22000-20629.00
; CURRENT APPLICATION NUMBER: US/09/847,102A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Homo sapiens
```


QY 69 -----FRWGTGLRMRCQQQLQDVSRRCRAAIRSMVRGYEEA 105
Db 411 XXXXXXXXXXXXXXXXXXXXWGQPPSRASRPV-----REVGRASARPLLGSSEP 466
QY 106 MEPLKGMWPGRQQQPPPGGGG-----QGGYY-----YPCSRPGGYGYG 148
Db 467 LCPASR-----PVQEGGGVSPPTPGQPRQGRGWGVSPPPGQP-SRPGRWG-G 513
QY 149 Q-----GGORQWYPCRP 161
Db 514 QPPPGQPHPGGEGRLYPASRP 535

RESULT 12
US-10-029-180-18
; Sequence 18, Application US/10029180
; Publication No. US20020182708A1
; GENERAL INFORMATION:
; APPLICANT: Cali, Brian M.
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin T.
; APPLICANT: Milna, G. Todd
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeffrey C.
; APPLICANT: Trueheart, Josh
; APPLICANT: Zhang, Lixin
; TITLE OF INVENTION: No. US20020182708A1el Regulators of Fungal Gene Expression
; FILE REFERENCE: MIC-004
; CURRENT APPLICATION NUMBER: US/10/029,180
; CURRENT FILING DATE: 2001-12-22
; PRIOR APPLICATION NUMBER: US 60/257,431
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fungal gene
US-10-029-180-18
Query Match 8.5%; Score 97.5; DB 14; Length 302;
Best Local Similarity 25.7%; Pred. No. 0.87;
Matches 49; Conservative 16; Mismatches 57; Indels 69; Gaps 10;
QY 26 EVERQL-----RLQCNQVQESPLDACRQVLDRLTGG-----GGG 63
Db 133 EYEQDEMSRANPPVHLTRELWC-AKMLGSIIDSSYDTKDERGSGSPGAPPPPGQYGY 191
QY 64 GGVGPRWGTGLRMRCQQQLQDVSRRCRAAIRSMVRGYEEAMPPLEKGMWPGRQQQPP 123
Db 192 GSGSP-GYSSAPYNGQQ-----QYSPAYPHSQA--P-GYAOQP 230
QY 124 PGGGGG-----QGGY-----YPCSRPGGYGYGQGGQRMYPPCR 161
Db 231 PGQYGYGYPGPGYSPAPYSSGQGGYGSAPYPPSSQASGYGQQYQGR---PGSQP 287
QY 162 GTTGGPRIGR 172
Db 288 GYPGQQPPYQG 298

RESULT 13
US-10-156-761-11120
; Sequence 11120, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11120
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11120
Query Match 8.5%; Score 97.5; DB 15; Length 623;
Best Local Similarity 38.1%; Pred. No. 1.8;
Matches 37; Conservative 6; Mismatches 33; Indels 21; Gaps 7;
QY 102 YEEAMPPLEKGM-WPWGRQQQPPP---QGG-GGGGGYYPYPCSRPGSGYGGCGGQROM 155
Db 76 YGQPPQGGQGGPGYGPQPPGYAQPPGYPGAPQPGYGPQQP--GYGYPQ-----127
QY 156 YPPCRPGT---TGGG---PRIGRVRLTKAREYAAAGL 185
Db 128 QPPQPFGAFTPPPGGSGKNPKGRPAVLIATAAVAL 164

RESULT 14
US-09-861-597-2
; Sequence 2, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861,597
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-861-597-2
Query Match 8.4%; Score 96.5; DB 9; Length 529;
Best Local Similarity 32.8%; Pred. No. 1.9;
Matches 41; Conservative 9; Mismatches 54; Indels 21; Gaps 5;
QY 55 RQLTGGGGGGVGPFRWGTGLRMRCQQQLQDVSRRCRAAIRSMVRGYEEAMPPLEKGMW 114
Db 45 QQPGGYGPGQGGPGRYGPG-----QQPGSGPAAAAAAGSGGQG--PGYGPGRQQGPG 97
QY 115 PWGRQQQPPPGG-----GGGGGGYYPYPCSRPGGYGGCGGQRMYPPCR 162
Db 98 GYGGQGGGSGPFGSAAAAAASAESGGPGGYGPGQGGPG--GYGPGQGGPGG--YGPQQG 155
QY 163 TTGGG 167
Db 156 PSGPG 160

```
RESULT 15
US-10-184-644-59
; Sequence 59, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 59
; LENGTH: 2331
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-59

Query Match      8.4%; Score 96; DB 15; Length 2331;
Best Local Similarity 25.6%; Pred. No. 9.1;
Matches 42; Conservative 2; Mismatches 56; Indels 64; Gaps 7;

Qy      5 AAAAAAALCPAALVAVAVCGEVEERLRLDLCWQEVQESPLDACRQVLDRLQTGGGGG 64
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      2080 AAAAAAACAACAAACAAACAA-----AAAAAGAC-----ATGGGAGA 2117

Qy      65 GVGPPRWGTG-LRMRCQQLQDYSRECRCAAIRSMVRYEEMPPLEKGMWPGRQQQPP 123
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      2118 GA---TGGGTGTTTTCC-----TCTCCACCACC----- 2142

Qy      124 PGGGGGGGGGYYPCSRPGEYGYGGGQRMYPFCRPGTTGGG 167
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      2143 -TGGGATA-----CGATGAGAGATGGCCACCTGCAAGCCAGG 2180
```

Search completed: November 29, 2003, 17:11:45

Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 29, 2003, 17:07:10 ; Search time 271 Seconds
(without alignments)
2051.971 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAALCFALVAV.....MMCRLEPQECIFSGGDQY 206

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DRV=xlh
-Q=/cg21_1/USPTO.spool/US10053410/runat_28112003_140756_27874/app_query.fasta_1.391
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -ENDS=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10053410 -CGN_1_1_0 @runat_28112003_140756_27874 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	281.5	24.5	2296	19	AAV20662	Barley D-hordein g
2	281.5	24.5	2380	22	AAF63380	Barley cDNA involv
3	281.5	24.5	2434	22	AAF63379	Barley cDNA involv
4	136	13.6	1302	24	AAI70852	Wheat glutenin gen
5	130.5	11.4	477	19	AAV20828	Nucleotide sequenc
6	129.5	11.3	109519	22	AAV20828	Micromonospora DNA
7	127.5	11.1	1089	23	ABL12131	Drosophila melanog
8	127.5	11.1	3089	23	ABL12130	Drosophila melanog
9	126.5	11.0	1847	22	AAH17970	Human cDNA sequenc
10	124.5	10.8	2042	24	ABN95855	Gene #2353 used to
11	123.5	10.8	4350	22	AAV20828	Human immunodefici
12	121.5	10.6	4347	21	AAA70724	Thermus thermophil
13	121	10.5	1029	25	ABZ22145	Thermus thermophil
14	121	10.5	2986	25	ABZ22145	Human protein Pp13
15	121	10.5	9968	22	AAI36353	Human musculoskele
16	121	10.5	9968	25	ABX59341	cDNA encoding nove
17	120.5	10.5	1806	22	AAH77885	Nucleotide sequenc
18	120.5	10.5	1806	24	ABK84228	Human cDNA differe
19	120.5	10.5	7954	22	AAK73498	Human immune/haema
20	120.5	10.5	21429	22	AAV57362	Murine Cdc25A comp
21	119.5	10.4	2794	22	AAV42090	Genomic sequence #
22	119.5	10.4	2794	22	ABL05837	Human reproductive
23	119.5	10.4	2794	23	ABL98401	Human testicular a
24	119	10.4	2456	22	AAV05659	Zea mays viviparou
25	119	10.4	2456	25	ABX13448	Zea mays viviparou
26	119	10.4	92638	24	ABQ88096	Human osteoblast d
27	118.5	10.3	5822	20	AAV62933	Human mdia Rho tar
28	118.5	10.3	5822	20	AAV62934	Human mdia Rho tar
29	118.5	10.3	9615	22	AAI36590	Human musculoskele
30	118.5	10.3	9615	25	ABX59578	cDNA encoding nove
31	117.5	10.2	1966	20	AAK61220	Mouse DNA demethyl
32	117.5	10.2	5123	24	ABV64378	Human von willebra
33	117	10.2	2429	23	ABL23142	Drosophila melanog
34	117	10.2	80557	24	ABX09142	Mycobacterium tube
35	117	10.2	4403765	22	AAI99683	Mycobacterium tube
36	117	10.2	4411529	22	AAI99682	Mycobacterium tube
37	116.5	10.1	3600	17	AAI13227	Thermotable enzym
38	116.5	10.1	5387	24	AAV39115	Human lung-specifi
39	116.5	10.1	10126	22	AAI99260	Human excretory re
40	116.5	10.1	10126	22	AAI63610	Human kidney relat
41	116	10.1	707	22	AAH34931	Human colon cancer
42	116	10.1	1233	22	AAH44351	Human nucleotide s
43	116	10.1	1233	22	AAH48220	Heart muscle cell
44	116	10.1	1233	22	AAH49586	Human coding sequ
45	116	10.1	1630	22	AAH16370	Human cDNA sequenc

ALIGNMENTS

RESULT 1

AAV20662

ID AAV20662 standard; DNA; 2296 BP.

XX AAV20662;

XX AAV20662;

DT 23-JUN-1998 (first entry)

XX Barley D-hordein gene 5'-terminal region.

DE DE Barley; D-hordein; Dpp3; promoter; gene expression; regulation;

XX seed; structural gene; ds.

KW Hordeum vulgare.

XX OS

XX WO9803655-A2.

XX

XX

XX

XX

XX

XX

XX

XX

XX

PD 29-JAN-1998. 104
 XX
 XX
 XX 22-JUL-1997; 97WO-JP02526. 366
 XX
 XX 23-JUL-1996; 96JP-0193433. 114
 XX
 XX (SAPB) SAPPORO BREWERIES. 426
 XX
 XX Hirota N, Ito K, Kihara M, Kuroda H; 120
 XX WPI; 1998-120779/11. 486
 XX
 XX Gene expression regulatory DNA, expression cassettes and vectors -
 PT comprising promoter region from barley, Hordeum vulgare, D-hordein
 PT gene, useful to control expression of desired gene e.g. to improve
 PT seeds
 PS
 PS Example 3; Page 25-26; 42pp; English. 133
 XX
 XX The present sequence represents the 5'-terminal region of the barley
 CC D-hordein gene. The present invention describes gene expression
 CC regulatory DNA which comprises a promoter region derived from the
 CC barley (Hordeum vulgare) D-hordein gene which allows expression of a
 CC desired structural gene, and a regulatory region regulating such
 CC expression. The introduction into plants of expression cassettes
 CC containing the gene expression regulatory DNA (either directly or via
 CC expression vectors) enables the expression of a gene within a plant cell
 CC e.g. barley to be controlled. The use of activating and suppressing
 CC regions in the regulatory DNA allows control of expression by e.g.
 CC tissue type or developmental stage, whilst the use of only an activating
 CC region maintains expression at a high level, providing an effective
 CC production means when recovery of the product of the structural gene is
 CC desired. The expression cassette/vector may be introduced into e.g.
 CC maturing seed endosperm tissue or regeneratable plant cells (e.g. derived
 CC from anthers) to improve seeds of barley or other plants, to produce
 CC gene products in seeds or to contribute to plant breeding programmes.
 CC The expression regulatory DNA can also be used in expression systems in
 CC vitro. GUS activity in barley protoplasts transfected with plasmid
 CC Dp3GUS2 comprising isolated D-hordein promoter region was 1.5 times
 CC higher than in those transfected with control pAct1f.
 XX
 XX Sequence 2296 BP; 629 A; 708 C; 606 G; 353 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 1.15e-08 Length: 2296
 Score: 281.50 Matches: 79
 Percent Similarity: 46.12% Conservative: 22
 Best Local Similarity: 36.07% Mismatches: 51
 Query Match: 24.52% Indels: 67
 DB: 19 Gaps: 11

US-10-053-410-4 (1-206) x AAV20662 (1-2296)

QY 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
 DB 37 ATGGCTAAGCGGTGGTCTCTTTGGCGGTATC-----GTCCGCTCTGGGTCTTC 90
 QY 21 AlaValCysGlnGlyGlyValGlu-----ArgGlnArgLeuArgAsp 34
 DB 91 ACCACCGCTGAACGCTGAGATCAATGGGAACAACATTTCCTGTAGCCGCTTAGGCAG 150
 QY 35 LeuGlnCysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAsp 54
 DB 151 CTACAGTGTGAGCGGAGCTCCAGGAGAGCTCGCTGAGCGGTGCGCGGGTCTGGAC 210
 QY 55 ArgGlnLeuThrGlyGlyGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGly 74
 DB 211 CAACAGCTG-----GTTGGCAGCTGCCATGGACACGGGG 246
 QY 75 LeuArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAla 94
 DB 247 CTCAGATGCACTGCTGCCAGCAGCTTCGGGAGCTCAGCCCGGAGTGGCGCCCGCTCGCC 306

QY 95 IleArgSerMetValArgGlyTyrGluGlu----- 104
 DB 307 CTCAGCCAGGTGCTGAGGCAATACGAGCAGAAACGAGGTGCCATCCAGGAGGATCC 366
 QY 105 -----AlaMetProProLeuGluLys---GlyTyrTrp 114
 DB 367 TTCTACCGCGGGGACCGCACCGCGCTGCAGCAAGGAGGATGGTGGGAACTCTGTGA 426
 QY 115 -----ProTyrGlyArgGlnGln----- 120
 DB 427 AAATGGTACTACCCAGACCAAACTTTCGCAACAGTCTATGGCAAGGGCAACAAGGGTAC 486
 QY 121 -----GlnProProGlnGlyGlyGlyGlyGlyGlnGlyGly 133
 DB 487 CACCAAGCGTAACCTTCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 546
 QY 134 TyrTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArg 153
 DB 547 TCAACTTTCCCG---CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGG 594
 QY 154 Gln-----MetTyrPro-----ProCysArgProGlyThrThrGlyGly 166
 DB 595 CAGCCATGCTCTATCCAAAGTCAACTTTCACCAACAGCAGCAGCAGCAGCAGGCAAGGG 651
 RESULT 2
 AAF63380
 ID AAF63380 standard; cDNA; 2380 BP.
 XX
 AC AAF63380;
 XX
 DT 10-MAY-2001 (first entry)
 XX
 DE Barley cDNA involved in D-hordein production SEQ ID 2.
 KW Barley; gel protein; D-hordein; malt; wort filterability; extraction; ss.
 OS Hordeum vulgare.
 XX WO200111946-A1.
 XX PD 22-FEB-2001.
 PF 16-AUG-2000; 2000WO-JP05476.
 PR 16-AUG-1999; 99JP-0229696.
 PA (SAPB) SAPPORO BREWERIES LTD.
 XX
 PI Hirota N, Kihara M, Ito K;
 XX WPI; 2001-191597/19.
 PT Transformation of barley with a D-hordein expression regulator for
 PT production of barley with lowered gel protein content -
 XX Claim 4; Page 33-34; 40pp; Japanese.
 CC This invention relates to barley having a low gel protein content. The
 CC barley is transformed with a polynucleotide sequence which regulated the
 CC formation of D-hordein. Transformation results in the production of
 CC barley strains with improved malting properties such as wort
 CC filterability and efficiency of extraction. The present sequence
 CC represents cDNA involved in the regulation of D-hordein production.
 SQ Sequence 2380 BP; 642 A; 747 C; 622 G; 369 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.18e-08 Length: 2380
 Score: 281.50 Matches: 79
 Percent Similarity: 46.12% Conservative: 22
 Best Local Similarity: 36.07% Mismatches: 51
 Query Match: 24.52% Indels: 67
 DB: 22 Gaps: 11

```

US-10-053-410-4 (1-206) x AAF63380 (1-2380)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
DB 24 ATGGCTAAGCGCTGGTCTCTTTGGCGGTAATC-----GTGCGCTCTGGGCTCTTC 77
QY 21 AlaValCysGlnGlyValGlu-----ArgGlnArgLeuArgAsp 34
DB 78 ACCACCGCTGAACGTCGAGATCAATGGAACAACATTTCTTGTATAGCGCTCTAGGCAG 137
QY 35 LeuGlnCysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAsp 54
DB 138 CTACAGTGTGAGCGCGAGCTCCAGGAGAGCTCGCTCGAGCGCTGCGCGGGTCTGGGAC 197
QY 55 ArgGlnLeuThrGlyGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGly 74
DB 198 CAACAGCTG-----GTTGGCCAGCTGCCATGGAGCGGGG 233
QY 75 LeuArgMetArgCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAla 94
DB 234 CTCCAGATGCACTGCTGCCAGCAGCTTCGGGAGCTCAGCCCGAGTGCCTCGCTCGCC 293
QY 95 IleArgSerMetValArgGlyTyrGluGlu-----104
DB 294 CTCAGCCAGTGTGAGGCAATACGAGCAGCAAAACCGAGGTGCCATCCAAAGGAGGATCC 353
QY 105 -----AlaMetProProLeuGluLys---GlyTrpTrp-----114
DB 354 TTCTACCCGGCGGAGCCCGCGCTGCAGCAAGGAGGATGGTGGGGAACCTCTGTGA 413
QY 115 -----ProTrpGlyArgGlnGln-----120
DB 414 AAATGGTACTACCCAGACCAAACTTTCGCAACAGTCAATGCGCAAGGGCAACAGGGTAC 473
QY 121 -----GlnProProProGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 133
DB 474 CACCAAGCGTAACCTTCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 533
QY 134 TyrTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArg 153
DB 534 TCAACTTTTCCCG---CAGCAGCAGCAGCAAGGACAA-----CAACGAGGACAGAG 581
QY 154 Gln-----MetTyrPro-----ProCysArgProGlyThrThrGlyGly 166
DB 582 CAGCCATGCTCTATCCAGTGCACACTTTCCCAACAGCAGCAGGCGCAAGGGCAAGGG 638

RESULT 3
AAF63379
ID AAF63379 standard; cDNA; 2434 BP.
XX
AC AAF63379;
XX
DT 10-MAY-2001 (first entry)
XX
DE Barley cDNA involved in D-hordein production SEQ ID 1.
XX
KW Barley; gel protein; D-hordein; malt; wort filterability; extraction; ss.
XX
OS Hordeum vulgare.
XX
PN WO200111946-A1.
XX
PD 22-FEB-2001.
XX
PF 16-AUG-2000; 2000WO-JP05476.
XX
PR 16-AUG-1999; 99JP-0229696.
XX
PA (SABP) SAPPORO BREWERIES LTD.
XX
PI Hirota N, Kihara M, Ito K;
XX

```

```

DR WPI; 2001-191587/19.
XX Transformation of barley with a D-hordein expression regulator for
XX production of barley with lowered gel protein content
XX Claim 4; Page 31-33; 40pp; Japanese.
XX This invention relates to barley having a low gel protein content. The
XX barley is transformed with a polynucleotide sequence which regulated the
XX formation of D-hordein. Transformation results in the production of
XX barley strains with improved malting properties such as wort
XX filterability and efficiency of extraction. The present sequence
XX represents cDNA involved in the regulation of D-hordein production.
XX Sequence 2434 BP; 675 A; 757 C; 626 G; 376 T; 0 other;

Alignment Scores:
Pred. No.: 1,218-08 Length: 2434
Score: 281.50 Matches: 79
Percent Similarity: 46.12% Conservative: 22
Best Local Similarity: 36.07% Mismatches: 51
Query Match: 24.52% Indels: 67
DB: 22 Gaps: 11

US-10-053-410-4 (1-206) x AAF63379 (1-2434)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
DB 37 ATGGCTAAGCGCTGGTCTCTTTGGCGGTAATC-----GTGCGCTCTGGGCTCTTC 90
QY 21 AlaValCysGlnGlyValGlu-----ArgGlnArgLeuArgAsp 34
DB 91 ACCACCGCTGAACGTCGAGATCAATGGAACAACATTTCTTGTATAGCGCTCTAGGCAG 150
QY 35 LeuGlnCysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAsp 54
DB 151 CTACAGTGTGAGCGCGAGCTCCAGGAGAGCTCGCTCGAGCGCTGCGCGGGTCTGGGAC 210
QY 55 ArgGlnLeuThrGlyGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGly 74
DB 211 CAACAGCTG-----GTTGGCCAGCTGCCATGGAGCAGCGGG 246
QY 75 LeuArgMetArgCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAla 94
DB 247 CTCCAGATGCACTGCTGCCAGCAGCTTCGGGAGCTCAGCCCGAGTGCCTCGCTCGCC 306
QY 95 IleArgSerMetValArgGlyTyrGluGlu-----104
DB 307 CTCAGCCAGTGTGAGGCAATACGAGCAGCAAAACCGAGGTGCCATCCAAAGGAGGATCC 366
QY 105 -----AlaMetProProLeuGluLys---GlyTrpTrp-----114
DB 367 TTCTACCCGGCGGAGCCCGCTGCAGCAAGGAGGATGGTGGGGAACCTCTGTGA 426
QY 115 -----ProTrpGlyArgGlnGln-----120
DB 427 AAATGGTACTACCCAGACCAAACTTTCGCAACAGTCAATGCGCAAGGGCAACAGGGTAC 486
QY 121 -----GlnProProProGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 133
DB 487 CACCAAGCGTAACCTTCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 546
QY 134 TyrTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArg 153
DB 547 TCAACTTTTCCCG---CAGCAGCAGCAGCAAGGACAA-----CAACGAGGACAGAG 594
QY 154 Gln-----MetTyrPro-----ProCysArgProGlyThrThrGlyGly 166
DB 595 CAGCCATGCTCTATCCAGTGCACACTTTCCCAACAGCAGCAGGCGCAAGGGCAAGGG 651

RESULT 4
AAI70852
ID AAI70852 standard; DNA; 1302 BP.

```


(USDA) US SEC OF AGRIC.

PA

XX Anderson OD;

PI

XX WPI; 1998-179218/16.

DR

XX P-PSDB; AAW53261.

DR

XX A process for altering the visco:elasticity of a dough - comprises
 PT genetic engineering of a high molecular weight glutenin subunit
 PT containing a non-natural repetitive domain

PT

PT

XX

PS Example 2; Page 17; 28pp; English.

XX

XX This nucleotide sequence is of the synthetic HMW-glutenin construct
 CC which is introduced into the DNA of the cereal to form a non-natural
 CC repetitive domain which changes the viscoelastic property of a dough.
 CC The domain comprises HMW glutenin subunit repeats which are synthetic
 CC in sequence or non-natural in number. The dough is useful for making
 CC a wheat flour, wheat dough, bread or noodle.

XX

SQ Sequence 477 BP; 91 A; 135 C; 134 G; 117 T; 0 other;

Alignment Scores:

Pred. No.:	4.43	Length:	477
Score:	130.50	Matches:	55
Percent Similarity:	37.62%	Conservative:	21
Best Local Similarity:	27.23%	Mismatches:	53
Query Match:	11.37%	Indels:	73
DB:	19	Gaps:	11

US-10-053-410-4 (1-206) x AAV20828 (1-477)

QY 24 GlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCys-----TrpGlnGlu 40

Db 4 GAAGGTGAAGCTGTGAACAG-----CTCAGTGCATCGCACTGCAGGAA 51

QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 60

Db 52 CTGCAGGACGTGAACCTGAAAGCTTCCAGCAGGTTATGAC----- 93

QY 61 GlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCys 80

Db 93 ----- 93

QY 81 GlnGlnGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100

Db 94 CAGCAGCTCGTGACATCTCTCCGGAATGCCACCCGGTGTGTTCTCCGGTTGCTGCG 153

QY 101 GlyTyrGluGlu-----AlaMetProProLeuGluLeuGlyGlyTyrTrpProTrpGlyArg 118

Db 154 CAGTACGACAGCAGATCGTTGTTCGCGC-----AAAGTGGTACCTTCTATCCGGT 207

QY 119 GlnGlnGlnProProGln-----GlyGlyGlyGlyGlyGlnGly 132

Db 208 GAACACCATCCCGCCGAGCAGCTGCAGCAGCTATCTTCTGGGGCATCCCGGCTCTGCTG 267

QY 133 GlyTyrTyrTyrPro-----CysSerArgProGlyGluGlyTyrGlyGln 149

Db 268 AAGCGCTACTACCCGCTGTACTTGTCTGCGCAG----- 300

QY 150 GlyGlyGlnArgGlnMetTyrPro-----ProCysArgProGlyThrThr--- 164

Db 301 -----CAGGTTCTTACTACCCGGTCTCAGGCTTCTCCGAGGTTCTAGTTCTTCTTAC 354

QY 165 -----GlyGlyGlyProArgIleGlyArgValArgLeuThrIysAlaArgGluTyrAla 182

Db 355 CACGTTCTGTGAACACACAGGCTGATCTCTGAAGTTGCTAAAGCTCAGCAGCTGGCT 414

QY 183 AlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePheSerGly 202

Db 415 GCGCAGCTCCCGCTATGTGCGGCTG-----GAAGGT 447

QY 203 GlyAsp 204

Db 448 GGTGAC 453

RESULT 6

AAS08693

ID AAS08693 standard; DNA; 109519 BP.

XX

AC AAS08693;

DT 26-SEP-2001 (first entry)

XX

DE Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.
 KW Everninomycin; antibiotic; bottle-neck gene; orthomycin;
 XX fermentation; ds.

OS Micromonospora carbonacea var. africana.

XX

FH Key Location/Qualifiers

CDS complement (132..1382)

FT /*tag= a

FT /product= "EvdA"

FT complement (1389..1394)

FT /*tag= b

FT complement (1490..2611)

FT /*tag= c

FT /product= "EvdB"

FT complement (2618..2622)

FT /*tag= d

FT complement (2622..3860)

FT /*tag= e

FT /product= "EvdC"

FT complement (3867..3870)

FT /*tag= f

FT 4143..5312

FT /*tag= g

FT /product= "EvdD"

FT 4134..4138

FT /*tag= h

FT 5309..6235

FT /*tag= i

FT /product= "EvdE"

FT 6232..7275

FT /*tag= j

FT /product= "EvdF"

FT 6226..6229

FT /*tag= k

FT 7272..8327

FT /*tag= l

FT /product= "EvdG"

FT 8342..9364

FT /*tag= m

FT /product= "EvdH"

FT 8333..8336

FT /*tag= n

FT complement (9463..10224)

FT /*tag= o

FT /product= "EvdI"

FT complement (10232..10235)

FT /*tag= p

FT 10424..11176

FT /*tag= q

FT /product= "EvdJ"

FT 12027..12455

FT /*tag= r

FT /product= "EvdK"

FT /partial

FT /note= "No start codon"

FT complement (12108..13022)

FT /*tag= s

FT /product= "EvdL"

FT complement (13027..13030)

FT /*tag= t

CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC in gene therapy. The primers sets can be used in antisense therapy and
 CC particularly full-length cDNAs. The primers are useful for synthesising polynucleotides,
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 1847 BP; 467 A; 505 C; 431 G; 444 T; 0 other;

Alignment Scores:
 Pred. No.: 27.4 Length: 1847
 Score: 126.50 Matches: 62
 Percent Similarity: 42.93% Conservative: 17
 Best Local Similarity: 33.70% Mismatches: 60
 Query Match: 11.02% Indels: 45
 DB: 22 Gaps: 9

US-10-053-410-4 (1-206) x AAH17970 (1-1847)

QY 18 ValAlaValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCys 37
 Db 1550 GTGGAGTGGAGGATGCCAGGCGCTCCCGGAAATGGAGAGGTGGAGGATTCAGGAC 1491
 QY 38 TrpGlnGluValGlnGlnSerProLeu-AspAlaCysArgGlnValLeuAspArgGlnLe 57
 Db 1490 CACCAGAGAGGGGAGGAGGAGGAGGTCCTCC-----TTCTCCAGGCAAGG 1440
 QY 57 uThr-GlyGlyGlyGly-----GlyGlyGly-ValGlyGlyProPhe 69
 Db 1439 GAGGAGTGGGGGGGAATTCACGCACTCCAGGCAAGAGGAGTGGTGGGGGATTCT 1380
 QY 70 -----ArgTrpGlyThrGlyLeuArgMetArgCysGlnGln 82
 Db 1379 AGCACTCCAGGCAAGAGGAGGTGGGGGATGGCAGTACTCTCCAGGCAAGAGAGG 1320
 QY 83 LeuGlnAspValSer-----ArgGluCysArgCysAlaAlaIleArgSer 97
 Db 1319 TGAAGGATGCCAACACCTCAGGCAAGAGGGGTGGAGGATGGTAGCATCCCCAGA 1260
 QY 98 MetValArgGlyTyr-GluGluAlaMetProProLeuGluLysGlyTrpTrpProTrpG 117
 Db 1259 CAAAGGAGGGGTGGAGAGATAGCAGTACTCTCCAGGTAAAGAGGG-----GG 1212
 QY 117 YArgGlnGlnProProPro-----GlnGlyGlyGlyGlyGlnGlyGlyTyrTy 136
 Db 1211 TGAGGAGATGCAACACCCCGAGGCAAGAGAGGTGGAGGAGGAGGAGGA----- 1160
 QY 136 rProCysSerArgProGlyGlyGlyTyrGlyTyr-----GlyGlnGlyGly 152
 Db 1159 -----GGAGGAGGAGGAGGAGTGGTACTATCCCGAGGAGCAGTGGTGG 1116
 QY 152 nArgGlnMetTyrPro-----ProCysArgProGlyThrThrGlyGlyGlyProArg 170
 Db 1115 TGAATAATAGTCCAGAGTCAACAGGTAAAGAGGGGCGAGGGGGAACAGGACGAGT 1056
 QY 170 egly 171
 Db 1055 AGGA 1052

RESULT 10

ABN95855/C

ID ABN95855 standard; DNA; 2042 BP.

XX

AC ABN95855;

XX

DT 13-AUG-2002 (first entry)

XX Gene #2353 used to diagnose liver cancer.
 DE
 XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KW metastatic liver tumour; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 XX Homo sapiens.
 OS
 XX WO200229103-A2.
 PN
 XX 11-APR-2002.
 PD
 XX 02-OCT-2001; 2001WO-US30589.
 PF
 XX 02-OCT-2000; 2000US-237054P.
 PR
 XX (GENE-) GENE LOGIC INC.
 PA
 XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 PI
 XX WPI; 2002-426119/45.
 DR
 XX
 XX Diagnosing and detecting the progression of liver cancer,
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,
 PT involves detecting the level of expression of two or more genes in a
 PT liver tissue sample
 PT
 XX Claim 1; SEQ ID NO 2353; 298pp; English.

XX The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN9503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2042 BP; 326 A; 755 C; 611 G; 350 T; 0 other;

Alignment Scores:
 Pred. No.: 39.9 Length: 2042
 Score: 124.50 Matches: 54
 Percent Similarity: 31.84% Conservative: 10
 Best Local Similarity: 26.87% Mismatches: 47
 Query Match: 10.84% Indels: 90
 DB: 24 Gaps: 11

US-10-053-410-4 (1-206) x ABN95855 (1-2042)

QY 48 AlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly-GlyGlyGlyGlyValG 67
 Db 899 GCGTCCCGAGTGGCGCCCGCG-----GGAGGAAGGAGGAGGCGGTGGGTGCG 849
 QY 67 YProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGlnLeuGlnAspVal 87
 Db 848 C-----AGGTGGGGT-----GGGGTC----- 833
 QY 87 rArgGluCysArgCysAlaAlaIleArgSerMetValArgGlyTyrGluGluAlaMetPr 107
 Db 832 -----TGGCGGAGGAGGAAGACCCACACCTGTGC 804
 QY 107 oProLeuGluLysGlyTrpTrpProTrpGlyArgGlnGlnProPro----- 123
 Db 803 TCCTCTCAGCAACTCCCGGGCGCCCTGGGGGCAACTCCCGTCCCGCCCGCCCATGGA 744


```
QY 83 LeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArgGlyTyr 102
Db 528 -----CCGCTACGGGACGAGATCGTGGCCCT 554
QY 103 GluGluAlaMetProProLeuGluLysGlyTyrTrpProTrpGlyArgGlnGlnPro 122
Db 555 GAGCGCGGCTTCCTCGGCGCTTCGCCCC---CATCTCGGGAGGTCCACGCCGCCCT 611
QY 123 ProProGlnGlyGly-----GlyGlyGlnGlyGlyTyrTyrProCys 138
Db 612 CGCGCCAGGAGCGGGGCTTCGTTGGAGGAGACCGCGGGGAGAGGGGTGCTCCGGGC 671
QY 139 SerArgProGlyGluGlyTyrGlyTyrGlyGlnArgGlnMetTyrProPro 158
Db 672 CCTCGAGCCAGCGGCGGAGCGGGAACGGGCGGACACCTGGTGGGGCCACCG 731
QY 159 CysArgProGlyThrThrGlyGlyGly-----ProArgIle 170
Db 732 GGACGACCTGCTCTCTCTCGAGGGCGCGCCGCCCGGTTGCGCAGCGCGGGGA 791
QY 171 GlyArg 172
Db 792 GCCCAA 797
RESULT 14
ID ABZ80413/c
XX ABZ80413 standard; cDNA; 2986 BP.
AC ABZ80413;
DT 28-MAY-2003 (first entry)
XX Human protein PP13671 encoding cDNA SEQ ID NO:18.
DE Human; cancer; cancer suppression; cancer inhibitor; gene; ss.
XX Homo sapiens.
PH Key Location/Qualifiers
FT CDS 1314..2414
FT /tag= a
FT /product= "Human protein PP13671"
XX
XX CN1368509-A.
XX
XX 11-SEP-2002.
XX
XX 08-FEB-2001; 2001CN-0105310.
XX
XX 08-FEB-2001; 2001CN-0105310.
XX
XX (SHAN-) SHANGHAI INST ONCOLOGY.
XX
XX Gu J;
XX
XX WPI: 2003-112778/11.
XX P-PSDB; ABP96556.
XX
XX Human protein that suppresses cancer cell growth and its coding
XX sequence -
XX
XX Claim 5; Page 24-26 (Disclosure); 36pp; Chinese.
XX
XX ABZ80408 to ABZ80418 encode the human proteins ABP96551 to ABP96561 which
XX have cancer inhibiting functions. Also described is a method for
XX preparing the proteins using recombination techniques. The human proteins
XX from the present invention, and nucleotide sequences encoding them, can
XX be used for treating diseases such as cancer.
XX
XX Sequence 2986 BP; 581 A; 976 C; 942 G; 487 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 92.9 Length: 2986
Score: 121.00 Matches: 59
Percent Similarity: 30.47% Conservative: 12
Best Local Similarity: 25.32% Mismatches: 62
Query Match: 10.54% Indels: 100
DB: 25 Gaps: 10
US-10-053-410-4 (1-206) x ABZ80413 (1-2986)
QY 5 AlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaValAlaValCysGln 24
Db 1292 GCCCTGCCGGCACCCCTCTCTGAGCTGGCCCTCCCTCGTC----- 1251
QY 25 GlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlnValGlnGlnSer 44
Db 1250 -----CTCCCTCCCTGGGGCGAAGACAGCAGTCG 1221
QY 45 ProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGlyGlyGlyGly 64
Db 1220 CCCTCCACCGTCTGTACTCCAGCCACC----- 1194
QY 65 GlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys-CysGlnGln----- 82
Db 1193 ---CTGCCTCTGTGAGATGGGA-----CGCTGCATGCTTAGAGGTCC 1152
QY 83 -----LeuGlnAspValSerArgGluCysArgCys-AlaAlaIleA 96
Db 1151 CTGTTCTCCCTACCCCTTTGCTGGGTCTCCGGCCCCACAGATGAGCACTGTC 1092
QY 96 rGserMetValArgGlyTyrGluGluAlaMetProProLeuGluLysGlyTyrTrpPro 116
Db 1091 GATATAAGCTC-----TGGCCCT 1074
QY 116 rPglyArgGlnGlnProProGlnGlyGly----- 127
Db 1073 GGCGGGCCAGCCTCGAGGACCGGCCAGTCTGGATGTTTCAGACCTGCGGGGCCCT 1014
QY 128 -----GlyGlyGlyGly 131
Db 1013 GGGGCTCTCCCTAGGAGACGAGGGCCACCTCTCTGCGGTTTCTTAGTCAGGACAA 954
QY 131 InGlyGlyTyrTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlyGly 151
Db 953 CAGGGGGCTCTGTAGAGCCA-----CGCGTGCAGGGGGCTGGGTACGGGTACA--- 904
QY 151 lYglnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyGlyProArgIle 171
Db 903 -----CCTCCCTCGGCCCTGGGACACCGAGTCTGCGGATACTTTGGG 861
QY 171 lYArgValArgLeuThrLysAlaArgGluTyrAlaAlaGlyLeuProMetCysArgL 191
Db 860 GTCCGGCAGCT-----GCAGGGGATATCCACCCCTGCA--- 826
QY 191 euSerGluProGlnGluCysSerIlePheSerGly 202
Db 825 --GAGGAGCCCGAGGAGCTCGGCGTCCGCGCTGGC 793
RESULT 15
AAL36353
ID AAL36353 standard; DNA; 9968 BP.
XX
XX AAL36353;
XX
XX 08-JAN-2002 (first entry)
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 2718.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX antineoplastic; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiatic; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein;
XX musculoskeletal system; ds.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 29, 2003, 18:16:16 ; Search time 311 Seconds
(without alignments)

2182.335 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAAALCFALVAV.....MMCRSLSEPQECIFSGGDQY 206

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh
-Q-/cgn2_1/USFTO_spool/US10053410/runat_28112003_140759_27990/app_query.fasta_1.391
-DB=Published Applications NA -QFMT=fastp -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10053410@cgn_1_1_271@runat_28112003_140759_27990
-NCPUS=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:
16: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:
17: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	127.5	11.1	9025608	14	US-10-156-761-1
					Sequence 1, Appli

C 2	123.5	10.8	4350	14	US-10-168-843A-33	Sequence 33, Appli
C 3	123	10.7	747	12	US-10-259-165-748	Sequence 748, App
C 4	122	10.6	930	14	US-10-156-761-4504	Sequence 4504, Ap
C 5	122	10.6	1578	14	US-10-156-761-3369	Sequence 3369, Ap
C 6	122	10.6	9025608	14	US-10-156-761-1	Sequence 1, Appli
C 7	121	10.5	1029	11	US-09-938-901-5	Sequence 5, Appli
C 8	121	10.5	9968	10	US-09-764-877-2718	Sequence 2718, Ap
C 9	120.5	10.5	1806	12	US-10-168-097A-1	Sequence 1, Appli
C 10	119.5	10.4	2794	11	US-09-764-891-8525	Sequence 8525, Ap
C 11	118.5	10.3	700	12	US-10-027-632-140092	Sequence 140092,
C 12	118.5	10.3	700	13	US-10-027-632-140092	Sequence 140092,
C 13	118.5	10.3	2988	12	US-10-027-632-111991	Sequence 111991,
C 14	118.5	10.3	2988	13	US-10-027-632-111991	Sequence 111991,
C 15	118.5	10.3	3031	12	US-10-027-632-115601	Sequence 115601,
C 16	118.5	10.3	3031	13	US-10-027-632-115601	Sequence 115601,
C 17	118.5	10.3	9615	10	US-09-764-877-2955	Sequence 2955, Ap
C 18	118	10.3	5862	12	US-10-132-134-15	Sequence 15, Appli
C 19	118	10.3	52101	12	US-10-132-134-1	Sequence 1, Appli
C 20	117.5	10.2	582	12	US-10-027-632-140201	Sequence 140201,
C 21	117.5	10.2	582	13	US-10-027-632-140201	Sequence 140201,
C 22	117.5	10.2	583	12	US-10-027-632-16086	Sequence 16086, A
C 23	117.5	10.2	583	13	US-10-027-632-16086	Sequence 16086, A
C 24	117.5	10.2	699	12	US-10-027-632-105402	Sequence 105402,
C 25	117.5	10.2	699	13	US-10-027-632-105402	Sequence 105402,
C 26	117	10.2	80557	15	US-10-080-170-647	Sequence 647, App
C 27	116.5	10.1	5387	13	US-10-001-873-22	Sequence 22, Appli
C 28	116	10.1	707	14	US-10-106-698-2023	Sequence 2023, Ap
C 29	116	10.1	1233	10	US-09-749-728B-2	Sequence 2, Appli
C 30	116	10.1	1329	14	US-10-156-761-6978	Sequence 6978, Ap
C 31	115.5	10.1	7727	12	US-10-204-925-5	Sequence 5, Appli
C 32	115	10.0	2152	12	US-09-873-319-712	Sequence 712, App
C 33	115	10.0	2152	12	US-09-960-706-1073	Sequence 1073, Ap
C 34	115	10.0	2432	9	US-09-925-301-574	Sequence 574, App
C 35	115	10.0	2432	12	US-10-262-473-5	Sequence 5, Appli
C 36	115	10.0	3148	14	US-10-084-817-219	Sequence 219, App
C 37	114.5	10.0	1299	14	US-10-156-761-4590	Sequence 4590, Ap
C 38	114	9.9	4167	9	US-09-764-878-282	Sequence 282, App
C 39	114	9.9	4167	9	US-09-764-860-1145	Sequence 1145, Ap
C 40	114	9.9	4167	10	US-09-764-846-345	Sequence 345, App
C 41	114	9.9	4167	11	US-09-764-891-7710	Sequence 7710, Ap
C 42	114	9.9	4167	11	US-09-764-891-10231	Sequence 10231, A
C 43	114	9.9	4167	12	US-10-212-872-1145	Sequence 1145, Ap
C 44	114	9.9	4167	14	US-10-091-483-345	Sequence 345, App
C 45	114	9.9	4167	14	US-10-079-854-282	Sequence 282, App

ALIGNMENTS

RESULT 1

US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (418715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 14.1 Length: 9025608
Score: 127.50 Matches: 42
Percent Similarity: 40.88% Conservative: 14
Best Local Similarity: 30.66% Mismatches: 45
Query Match: 11.11% Indels: 36
DB: 14 Gaps: 7
US-10-053-410-4 (1-206) x US-10-156-761-1 (1-9025608)
QY 59 GlyGlyGlyGlyGlyGlyValGlyValGlyThrGlyLeuArgMetArg 78
Db 4272268 GGTTCGAACCTGGGGCGGCATATGACCGGGCGCTGGCCACCGGGACTCCCGCGCGC 4272209
QY 79 CysCysGlnLeuGlnAspValSerArgGluCys-----ArgCysAlaAlaIleArg 96
Db 4272208 TGCTGCGCGCTTTCGCG-----TGCTGGCGCGGTGCTGCTGAGCTGCGG 4272164
QY 97 SerMetValArgGlyTyrGluGluAlaMetProLeuGluLysGlyTyr-----113
Db 4272163 GGCCTGGGGCTGGGTGG-----CGGGGGTGGTGGCGGTGG 4272128
QY 114 -----TrrProTrrpGlyArgGlnGlnProProGlnGly-Gly-- 127
Db 4272127 CGCGGGGGTGGCGGTGGGTGGTTCACGCGGGAGCGCGTGGCGGAGCGC 4272068
QY 128 -----GlyGlyGlnGlyGlyTyrTyrTyrProCysSerAr 140
Db 4272067 GCTGTGCGGGCGCGACAGCGCGGTGGCGGGGTGGCGAGCTGC---TGTGCGGG 4272011
QY 140 gProGlyGluGlyTyrGlyGlnGlyGly-----GlnArgGlnMetTyrPr 157
Db 4272010 GCCGTTCCTGTGATCTGGCACCGCTGGCGGGCGCGCTCAGGGGTGGCACGCTGG 4271951
QY 157 oProCysArgProGlyThrThrGlyGlyProArgIleGlyArgVal 173
Db 4271950 CCCTGGTTCGGCGGTGGGTGGCGAGTGGCGGTGCCAGGTTCGGGTG 4271902
RESULT 2
US-10-168-843A-33/c
; Sequence 33, Application US/10168843A
; Publication No. US20030108562A1
; GENERAL INFORMATION:
; APPLICANT: Medical Research Council
; APPLICANT: International Aids Vaccine Initiative
; APPLICANT: University of Nairobi
; TITLE OF INVENTION: Improvements in or Relating to Immune Responses to HIV
; FILE REFERENCE: MJL/C1248/1/M
; CURRENT APPLICATION NUMBER: US/10/168,843A
; CURRENT FILING DATE: 2002-09-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 4350
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: polynucleotide
US-10-168-843A-33
Alignment Scores:
Pred. No.: 0.0229 Length: 4350
Score: 123.50 Matches: 42
Percent Similarity: 38.89% Conservative: 7
Best Local Similarity: 33.33% Mismatches: 54
Query Match: 10.76% Indels: 23
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (418715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-053-410-4 (1-206) x US-10-168-843A-33 (1-4350)
QY 55 ArgGlnLeuThrGlyGlyGly-----GlyGlyGlyValGlyProPheArgTrpGly 72
Db 2296 CGGCGCGCCAGCGCGCGCGGCTTCTTGGCGGCGCTGGCGGAGCCCTTGGGCTTGGGT 2237
QY 73 ThrGlyLeuArgMetArg-----CysCysGlnGlnLeuGlnAspValSerArg 88
Db 2236 AGGGGTGGACGTGTCACAGATCTTGTATGATCGATGCGCTTTCAGCAGCTCTCTCGTCGG 2177
QY 89 GluCysArgCysAlaAlaIleArgSerMetValArgGlyTyrGluGluAlaMetProPro 108
Db 2176 AGTTGCGGAGCGCGC-----CGGCGTCTCGGGGTGGTCTCGGTCTTGGACTCCA 2126
QY 109 LeuGluLysGlyTrrProTrrpGlyArgGlnGlnProProGlnGlyGly 128
Db 2125 ---CCTTCTTCTTGGACTCTTGGGCGGTGGAGATGCCCT-----GGGTCTGGG 2078
QY 129 Gly-----GlyGlnGlyGlyTyrTyrTyrProCysSerArgProGly 142
Db 2077 GATGGGTCTGTTGGGATGGGTCTTGTGTGT-----CCTTGTGGACTGGGGG 2027
QY 143 ---GluGlyTyrGlyTrrpGlyGlnGlnArgGlnMetTyrProCysArgPro 161
Db 2026 TGCGCGCGCGTGGCGCGCTTCTTGGCGCGGTAGAGATGCCAGGCCCTTGTTCAGGA 1967
QY 162 GlyThrGlyGlyGly 167
Db 1966 AGCACACGCGGCGAGTGGT 1949
RESULT 3
US-10-259-165-748
; Sequence 748, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Rieke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 748
; LENGTH: 747
; TYPE: DNA
; ORGANISM: Zea mays
US-10-259-165-748
Alignment Scores:
Pred. No.: 0.00475 Length: 747
Score: 123.00 Matches: 43
Percent Similarity: 31.45% Conservative: 7
Best Local Similarity: 27.04% Mismatches: 47
Query Match: 10.71% Indels: 62
```


Alignment Scores:

Pred. No.: 0.01 Length: 1029
Score: 121.00 Matches: 41
Percent Similarity: 39.34% Conservative: 7
Best Local Similarity: 33.61% Mismatches: 44
Query Match: 10.54% Indels: 30
DB: 11 Gaps: 4

US-10-053-410-4 (1-206) x US-09-938-901-5 (1-1029)

QY 63 GlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGln 82
DB 486 GGCCTGGGGGGAGGGCTTTCCTGGGAGCGGAGCTGCG-----527
QY 83 LeuGlnAspValSerArgGluCysArgCysAlaAlaileArgSerMetValArgGlyTyr 102
DB 528 -----CCGCTACGGGAGCAGATCGTGGCCCT 554
QY 103 GluGlnAlaMetProLeuGluLysGlyTyrTrpTrpGlyArgGlnGlnPro 122
DB 555 GAGGCGCGCTTCTCGGGCTTCGCCC---CATCTCGGGAGGTCCACGCGCCCT 611
QY 123 ProProGlnGlyGly-----GlyGlyGlnGlyGlyTyrTyrProCys 138
DB 612 CGCCGCAAGGAGCGGGCTTCGCTTGGAGGAGACCGGGGGAAGGGTGTCTCGGGC 671
QY 139 SerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProPro 158
DB 672 CTTGAGGCGCAGCGGGCGGAGAGCGGGAACGGGCCAGACCTCGTGGGGCCCCCACC 731
QY 159 CysArgProGlyThrThrGlyGlyGly-----ProArgIle 170
DB 732 GAGGACCTGTCTTCTCTGAGGGGGCGGCGCCCGCCCGCTTCGCGAGCGCGGGGA 791
QY 171 GlyArg 172
DB 792 GGCCTAA 797

RESULT 8

US-09-764-877-2718
; Sequence 2718, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2718
; LENGTH: 9968
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2718

Alignment Scores:

Pred. No.: 0.0876 Length: 9968
Score: 121.00 Matches: 60
Percent Similarity: 37.57% Conservative: 11
Best Local Similarity: 31.75% Mismatches: 65
Query Match: 10.54% Indels: 54
DB: 10 Gaps: 10

US-10-053-410-4 (1-206) x US-09-764-877-2718 (1-9968)

QY 5 AlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaValCysGln 24
DB 89 GCGGCGCCGTGTACCCGCTGTCTCGCTCGCGCTCGCGCCGACGCGCCCTGCCTG 148
QY 25 GlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlnValGlnGlnSer 44

DB 149 GGCACATCGACTTCAGGAGCTGCGC---ATGATGCTGCTAAGCGAGCGCGCGCGCG 205
QY 45 ProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrClyGlyGlyGlyGly 64
DB 206 -----GAGAAGACGTGCGGCGCGCGCGCGCGGA 232
QY 65 GlyValGlyProPheArg-----TrpGlyThrGlyLeuArgMetArgCysCys 80
DB 233 GCGAG-GGCGCGGGCGGCGCGCCCAAGGAGGCGCGGG-----273
QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaileArgSerMetValArg 100
DB 274 -----CGCGCTCGCGCTCGCGCTGCGACCGACGCGCGCGCGCGCG 312
QY 101 GlyTyrGluGluAla-----MetProPro-LeuGluLysGlyTyr-----113
DB 313 GGCCTGGGCGCAGCGCGCGCGCTGCGCTCGGAGGACCGCTTCAACAGCATCCCGCG 372
QY 114 -----TrpProTrpGlyArgGlnGlnProProGlnGlnGlyGly-----128
DB 373 CGAGCA 432
QY 129 -----GlyGlyGlnGly-----GlyTyrTyrTyrProCysSerArgPr 141
DB 433 GTGCGTGCCTGCG 483
QY 141 GglyGluGlyTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProPro---CysAr 160
DB 484 GGGCG 543
QY 160 gProGlyThrThrGlyGlyGlyPro 168
DB 544 TCCGCGCGCACAGCGCGCGGACCT 568

RESULT 9

US-10-168-097A-1/c
; Sequence 1, Application US/10168097A
; Publication No. US20030166245A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL OF THE RECHERCHE SCIENTIFIQUE
; APPLICANT: INSTITUT CURIE
; TITLE OF INVENTION: WASP FAMILY PROTEIN FRAGMENTS, AND USES THEREOF
; FILE REFERENCE: IFB99WASP
; CURRENT APPLICATION NUMBER: US/10/168,097A
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1806
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (35)..(1540)
US-10-168-097A-1

Alignment Scores:

Pred. No.: 0.0191 Length: 1806
Score: 120.50 Matches: 39
Percent Similarity: 35.14% Conservative: 0
Best Local Similarity: 35.14% Mismatches: 22
Query Match: 10.50% Indels: 51
DB: 12 Gaps: 5

US-10-053-410-4 (1-206) x US-10-168-097A-1 (1-1806)

QY 60 GlyGlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys 79
DB 1284 GGAGGAGTGGGGAGGGGCTGCTCATTC---GGAGCTGGCGCGCGGTGGCGG---1232
QY 80 CysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaileArgSerMetVal 99

```
Db 1232 ----- 1232
QY 100 ArgGlyTyrGluGluAlaMetProProLeuGluLysGlyTrpTrpProTrpGlyArgGln 119
Db 1231 -----TGGTGGCGGTGGTGGTGGCAT 1211
QY 119 nGlnGlnProProGlnGlnGlyGlyGlyGlnGlyGlyTyrTyrProCys 139
Db 1210 GGTGGCCACAGCTCCAGGGGGTGGAGGGGAGTGGTCCA-----GA 1166
QY 139 rArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArgGlnMetTyrProProCys 159
Db 1165 ACCTCCAGTAGCTGGAGGGGGTGGTGGAGGG-----CCCCC 1127
QY 159 sArgProGlyThrThrGlyGlyGlyProArg 169
Db 1126 TCGGCCTGGG-----GGTGGGGGTCCCCGG 1102

RESULT 10
US-09-764-891-8525
; Sequence 8525, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8525
; LENGTH: 2794
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8525

Alignment Scores:
Pred. No.: 0.0361 Length: 2794
Score: 119.50 Matches: 40
Percent Similarity: 38.76% Conservatives: 10
Best Local Similarity: 31.01% Mismatches: 49
Query Match: 10.41% Indels: 30
DB: 11 Gaps: 6

US-10-053-410-4 (1-206) x US-09-764-891-8525 (1-2794)
QY 60 GlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys 79
Db 103 GGTGGGGTGGGGCGGACCTGGCCCTGGTGTGGCCCTCACCTTTCTCCACACA 162
QY 80 CysGlnGlnLeuGlnAsp-----ValSerArgGluCysArgCysAlaAla 95
Db 163 GGCCAGAACTACATGAAGGGCTCCCGGAATTGGAGAAGGATTTTCCTC----- 216
QY 96 ArgSerMetValArgGlyTyrGluGluAlaMetProProLeuGluLysGlyTrpTrpPro 115
Db 217 -----TATCTGACCAATGAAGCCCTCTGGGTACGGCTGGCTAGTG 258
QY 116 TrpGlyArgGlnGlnProProGlnGlyGlyGlyGlyGlyGlnGly----- 132
Db 259 TGGGGG-----CACCGGTGGTGGAGTGGGGGCTCCGGTGTGGCCTG 306
QY 133 -----GlyTyrTyrProCysSerArgProGlyGluGlyTyrGlyTrpGlyGlnGly-- 150
Db 307 ATGCACCTTCCCGCAGCTGTGAACCTCTCGAGAAGATGCTGGTGTGGACGCGAGC 366
QY 151 -----GlyGlnArgGlnMetTyrPro-----ProCysArgProGln 162
Db 367 AGCGGTGACGGAGCGAGCGCTGGCCCATCTTCTGAGTCCCTGCACGACACGG 426
QY 162 YThrThrGlyGlyProArgIle 170
|||||
|||
```

```
Db 427 AAGATGAGCCCGGAGGTCCAGAGTA 451

RESULT 11
US-10-027-632-140092/c
; Sequence 140092, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140092
; LENGTH: 700
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140092

Alignment Scores:
Pred. No.: 0.012 Length: 700
Score: 118.50 Matches: 54
Percent Similarity: 37.78% Conservatives: 14
Best Local Similarity: 30.00% Mismatches: 49
Query Match: 12.32% Indels: 63
DB: 12 Gaps: 10

US-10-053-410-4 (1-206) x US-10-027-632-140092 (1-700)
QY 58 ThrGlyGlyGlyGlyGly----- 64
Db 504 ACCGGCGGGCGGAGCGCGGCTCGGGGGGAGCTCGGGCGGCGGCACGTCA 445
QY 65 GlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCysGlnGlnLeuGln 84
Db 444 GGGGGGGGGCGCGCGCGGGAAGT-----CTCTGTAGGGCGGCGGC 400
QY 84 nAspValSerArgGluCys-----ArgCysAlaAlaAlaAlaArgSerMetVal 100
Db 399 TACCCTCAGCCCGCGCGCTCATGGCCCTTTCGGTCCCGGCTACTCACCGGGCTTCCGA 340
QY 100 rGlyTyrGluGluAlaMetProProLeuGluLysGlyTrpTrpProTrpGlyArgGlnG 120
Db 339 AAGCCCGCGGAGTAGTGGCTCCGACGAAAGGCGCGGAGCGGTGGAGCTGGCGCC 280
QY 120 LngLngProProGlnGlyGlyGlyGlyGlyGlnGlyGlyTyrTyrProCysSerA 140
Db 279 TCCCCCGCGGTGAGCTGACGAGCGCGGCG-----CCCGCGAGCC 238
QY 140 rGProGlyGluGlyTyrGlyTyrGlyGln-----GlyGlyGlnArgGln----- 154
Db 237 GGCCTGGTGGGGGGTGGCTCTTCGCCCTTTCCTGTGCGGGGGGCGAGAGCGGTGGC 178
QY 155 -----MetTyrProProCysArgProGlyT 163
Db 177 AGCGGGCGGCGCGCGCGCTGCTCGAGGAGAACCCCTTCGCGCGCTGGACACCG---A 121
```


US-10-053-410-4 (1-206) x US-10-027-632-115601 (1-3031)

```
QY 44 SerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGlyGlyGly 63
Db 2904 AGTCCCTCTCAGCC-----CAGCTGCCAGTGGAGGGGTGAG 2866
QY 64 GlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGlnLeu 83
Db 2865 GCGCGTCTGGGA-----GATACGGCC 2845
QY 84 GlnAspValSerArgGluCys-----ArgCysAla-----AlaIleArgSerMet 98
Db 2844 CAGGAGAGAGAGCTCCCTGCTGACCGGCTGTGCCCGACAGCTCGGCCATCGACAAGCG 2785
QY 99 ValArg-----GlyTyrGluGluAlaMetProProLeu----- 109
Db 2784 GTGAGGGAGAGAGTGGCCATCAGAGAGCTGAGCCGACCCCTTTCAGTCCGAGATCTTCGC 2725
QY 110 -----GluLysGlyTrp 113
Db 2724 CAAGCGCGCTACCGGGAGCTGCTGCTGAAGCACATGCAGCATGAGAAGCTAGGTGG 2665
QY 114 TrpProTrpGlyArgGlnGlnGlnProProProGlnGlyGlyGlyGlnGlyGly 133
Db 2664 TGG-----CTGCCCCGGGAGAGGGGTGGGGCTGCCCTGGG 2629
QY 134 TyrTyrTyrProCysSerArgProGlyGluGlyTyrGly-----TyrGlyGln 149
Db 2628 -----CAGTCAGGGGAGGGCGGAGGGGGCTCCCGGTATGGAGAG 2587
QY 150 -GlyGlyGlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyProAr 169
Db 2586 CTCGGGACAGTCT-----CCTCCCTGCTCCCTGAC----- 2556
QY 169 gIleGlyArgValArgLeuThrLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCy 189
Db 2555 -----GGTGCCTCTCGCCTCACTATTGAAAGAGGGGGTGGCGAGACCTTCCC-----TG 2506
QY 189 sArgLeuSerGluProGlnGluCysSerIlePhe 200
Db 2505 CCACATTTCCAGGAGAGGTCAAGGGTATTTT 2472
```

Search completed: November 29, 2003, 21:28:17
Job time : 6283 secs

THIS PAGE BLANK (USPTO)